CALIFORNIA INSTITUTE OF TECHNOLOGY

EARTHQUAKE ENGINEERING RESEARCH LABORATORY

Application of Stochastic Simulation Methods to System Identification

BY

MATTHEW M. MUTO

REPORT NO. EERL 2007-02

PASADENA, CALIFORNIA

MAY 2007

A REPORT ON RESEARCH SUPPORTED BY THE CALIFORNIA INSTITUTE OF TECHNOLOGY UNDER THE SUPERVISION OF JAMES L. BECK.

Application of Stochastic Simulation Methods to System Identification

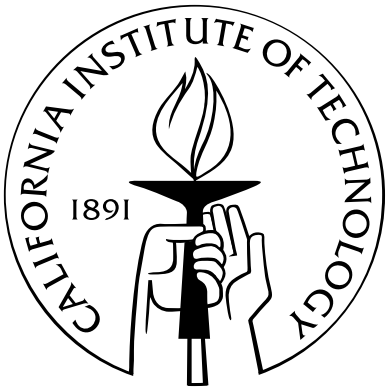
Thesis by

Matthew M. Muto

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2006

(Submitted May 21, 2007)

Acknowledgements

I would first like to gratefully acknowledge the efforts of my advisor, Dr. Jim Beck, whose guidance, encouragement, insight, and patience made this thesis possible. I would also like to thank Dr. John Hall and Dr. Tom Heaton for their contributions to my growth as a scientist and an engineer during my years here at Caltech. I am also grateful to my undergraduate advisor, Dr. Zee Duron, for his confidence in me, which helped to set me on this path and to keep to it in the face of adversity.

I benefitted greatly from my collabration with Dr. Jianye Ching and owe much of my understanding of stochastic simulation methods to him. I would also like to thank him for graciously allowing me the use of his transitional Markov Chain Monte Carlo algorithm prior to its publication.

I owe a great deal to my friends, inside and outside Caltech. I have learned from them, and their friendship has helped me through the highs and lows of grad school. I would like to thank John, Andy, Case, Charlotte, Swami, Steve, Steve, Jeff, Judy, Alex, Bayani, Michelle, and the many, many others who have made my time here so special.

I would also like to thank Carolina Oseguera and the rest of the Thomas building staff for their dedicated efforts on behalf of myself and all the other graduate students here.

Finally, and most importantly, I must acknowledge the love and support of my family: my parents, Paul and Myrna; my brother, Adam; my sisters, Gina and Maria; my grandparents, aunts, uncles, and cousins everywhere. I could not have accomplished this without you.

Abstract

Reliable predictive models for the response of structures are a necessity for many branches of earthquake engineering, such as design, structural control, and structural health monitoring. However, the process of choosing an appropriate class of models to describe a system, known as model-class selection, and identifying the specific predictive model based on available data, known as system identification, is difficult. Variability in material properties, complex constitutive behavior, uncertainty in the excitations caused by earthquakes, and limited constraining information (relatively few channels of data, compared to the number of parameters needed for a useful predictive model) make system identification an ill-conditioned problem. In addition, model-class selection is not trivial, as it involves balancing predictive power with simplicity.

These problems of system identification and model-class selection may be addressed using a Bayesian probabilistic framework that provides a rational, transparent method for combining prior knowledge of a system with measured data and for choosing between competing model classes. The probabilistic framework also allows for explicit quantification of the uncertainties associated with modeling a system. The essential idea is to use probability logic and Bayes' Theorem to give a measure of plausibility for a model or class of models that is updated with available data.

Similar approaches have been used in the field of system identification, but many currently used methods for Bayesian updating focus on the model defined by the set of most plausible parameter values. The challenge for these approaches (referred to as asymptotic-approximation-based methods) is when one must deal with ill-conditioned problems, where there may be many models with high plausibility, rather than a single

dominant model. It is demonstrated here that ill-conditioned problems in system identification and model-class selection can be effectively addressed using stochastic simulation methods.

This work focuses on the application of stochastic simulation to updating and comparing model classes in problems of: (1) development of empirical ground motion attenuation relations, (2) structural model updating using incomplete modal data for the purposes of structural health monitoring, and (3) identification of hysteretic structural models, including degrading models, from seismic structural response.

The results for system identification and model-class selection in this work fall into three categories. First, in cases where the existing asymptotic approximation-based methods are appropriate (i.e., well-conditioned problems with one highest-plausibility model), the results obtained using stochastic simulation show good agreement with results from asymptotic-approximation-based methods. Second, for cases involving ill-conditioned problems based on simulated data, stochastic simulation methods are successfully applied to obtain results in a situation where the use of asymptotics is not feasible (specfically, the identification of hysteretic models). Third, preliminary studies using stochastic simulation to identify a deteriorating hysteretic model with relatively sparse real data from a structure damaged in the 1994 Northridge earthquake show that the high-plausibility models demonstrate behavior consistent with the observed damage, indicating that there is promise in applying these methods to ill-conditioned problems in the real world.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Characterizing and, perhaps more importantly, predicting the response of civil structures under extreme loading events such as earthquakes is a challenging problem. The scale of structures such as buildings, bridges, and dams, and their interaction with the surrounding earth, precludes the study of complete systems in a laboratory setting, and it is difficult to produce the response amplitiudes caused by large seismic events in field tests. Variability in the material properties, quality of construction, environmental conditions, and complex constitutive behavior of structural materials make it impossible, or at least infeasible, to construct reliable deterministic structural models from first principles. One approach to this problem is to use the recorded seismic response from an instrumented structure to extract the parameters of a numerical model of the structure (Pilkey and Cohen, 1972; Hart and Yao, 1976; Beck, 1978, 1996), a process known as *system identification*. However, even with the extensive instrumentation available in many seismically active regions, system identification can be difficult to carry out, since it is often an ill-conditioned inverse problem. Another issue in system identification is the selection of the *model class* to be identified, where identifying a more complicated model must be weighed against its usefulness for making accurate predictions. There may also be considerable uncertainty in the identified model parameters, due to the sources of variability mentioned earlier, measurement error, and the inherent inability of a mathematical model to exactly capture the behavior of a real structure.

The challenges associated with the problem of system identification in earthquake

engineering may be addressed through the use of a *Bayesian probabilistic framework*, in which probability is treated as a multi-valued logic that may be used to perform plausible inference. A Bayesian approach is capable of performing identification of ill-c onditioned systems as well as providing a basis for a rational quantitative comparison of model classes, while explicitly including the effects of the uncertainty in the problem. However, many current techniques for implementation of this framework cannot be effectively implemented for ill-conditioned systems with many parameters.

This dissertation presents the application of techniques known as *stochastic simulation methods* to system identification in civil engineering problems, in particular to the updating of structural models using recorded seismic response, working within a Bayesian probabilistic framework.

1.1 System Identification

Typically, the approach to system identification is viewed as the process of identifying a model for a physical system using available data (Eykhoff, 1974; Goodwin and Payne, 1977; Doebling et al., 1996; Ljung, 1999; Nelles, 2001). This work focuses on *parametric* system identification, where a model within a chosen model class is defined by assigning values to a set of model parameters. Identification is usually performed by finding the "best" set of parameters for the given model class, according to a chosen criterion such as least-squares error. If the resulting model is deemed to be adequate for its intended application, predictions are made using this optimal model; if not, the model class is revised, and the process is repeated. Successful implementation of this method often requires the user to make difficult judgements without a clearly defined basis for making comparisons and assessing performance. Nonetheless, there is a high level of sophistication and effectiveness in current system-identification methods. However, solutions are often somewhat ad hoc and require fine-tuning for each problem. Users of system-identification methods must have considerable experience to avoid under- or over-fitting of the data, or to deal with ill-conditioned inverse problems with multiple optimal (or near optimal) solutions. In short, there is no

unifying framework to guide the user in selecting the model class, determining the criteria used to compare models, etc.

An alternative approach, which has gained a substantial following in many research disciplines in recent years, is to evaluate candidate models using the probability of the models given data from the system of interest. However, probability is used not in the classical sense as the long-run frequency of some event, but rather as a multi-valued logic that expresses the degree of plausibility of a given proposition, such as the value of a model parameter being contained in a given range, or the failure of a structure under earthquake loading during its expected lifetime.

The roots of the probability logic approach are in work performed by Reverend Thomas Bayes and published posthumously (Bayes, 1763, 1958). He presented a method for updating probability distributions for parameters based on available data that would come to be known as *Bayes' Theorem*, and it forms the foundation of a framework for probabilistic inference. It was Pierre Simon Laplace (1781, 1812, 1951), however, who showed the power of Bayes' Theorem by applying it to problems using real data and demonstrating the ability of probabilistic inference to separate "signal" from "noise." Other pioneers in Bayesian probabilistic inference and its applications to real problems are Harold Jeffreys (1939) and Richard Cox (1961), who developed formal axioms for probable inference. The culmination of this work is the theory of probability as multi-valued logic as presented by Jaynes (1957, 2003).

The probability logic approach is applied to system-identification problems using the probabilistic framework developed by Beck (1989, 1996) and Beck and Katafygiotis (1991, 1998). This framework has been successfully applied to a wide variety of engineering systems, such as model updating for reliability (Beck and Au, 2002) and structural health monitoring (Vanik et al., 2000; Yuen et al., 2004). This framework has also been applied to more challenging problems such as updating non-linear systems with uncertain input (Yuen and Beck, 2003b) and reliability-based control robust to probabilistic model uncertainty (May and Beck, 1998; Yuen and Beck, 2003a; Scruggs et al., 2006).

When addressing the problem of system identification, the probability logic ap-

proach does not necessarily demand starting from the ground up. It can be shown that many effective methods in system identification can be equivalently expressed in the context of probability logic. However, in the probability logic framework, any assumptions and simplifications in a particular problem are explicitly stated as the conditioning information in the probability statements, rather than being implicitly incorporated into the identification method. It is therefore easier to understand the various advantages and limitations associated with these assumptions. This method also allows for the uncertainties associated with model parameters to be quantified, which can be important in engineering applications.

1.2 Model-Class Selection

Once it has been decided how system identification will be performed on a model class, the question still remains: What model class should be used? There may be many potential model classes considered by the user for a given process. *Model-class selection* is the process of choosing between competing model classes based on available data. It can be viewed as a generalization of system identification to the model-class level, as it is usually desired to find an "optimal" model class among the pool of candidate model classes. The ability of a model class to fit the available data is an important consideration. However, a more complicated model can typically fit the data better than a less complicated one with fewer parameters. If the optimal model class is chosen purely by minimization of the error between the data and the corresponding predictions of the optimal model in each class, the optimal model class will be biased in favor of more complicated models. This approach is therefore likely to lead to over-fitting of the data. When an over-fitted model is used to make predictions, it typically yields poor results because the model will be overly dependent on the details of the data, giving predictions that reflect characteristics of the noise in the data set as well as the behavior of the system of interest. This point was noted for structural mechanics problems by Grigorou et al. (1979).

The need for considering both the quality of the data fit and the model complexity

when choosing between model classes was recognized earlier by Jeffreys (1939). This work pointed out a need for a "simplicity postulate" that would give a quantitative expression of the principle known as "Ockham's razor": simpler models that are consistent with the data should be preferred over more complex models that offer only slight improvements in the fit to the data. In a non-Bayesian context, Box and Jenkins (1970) expressed the same idea in citing the importance of developing parsimonious models for time-series forecasting.

Early efforts at quantifying this desire for parsimony used ad hoc penalty terms against over-parameterization (Akaike, 1974). However, such methods require considerable experience and fine-tuning to be effective in selecting model classes for use in system identification. The probability logic approach can easily be extended to model-class selection and has been shown to automatically enforce model parsimony (Schwartz, 1978; Gull, 1989; Sivia, 1996; Beck and Yuen, 2004). The extension of the probabilistic framework to model-class selection allows for consistent, rational comparisons of competing model classes based on the available data and conditioning information.

1.3 Stochastic Simulation Methods

While the power of the probability logic approach to system identification and model-class selection has been well demonstrated, there are still computational issues to be resolved in its implementation, particularly when dealing with ill-conditioned problems. Stochastic simulation methods are a class of techniques that have been applied to perform Bayesian inference (Geman and Geman, 1984; Gelfand et al., 1990; Gilks et al., 1996) and also directly applied to system identification under the Bayesian framework (Beck and Au, 2002). The methods have more recently been applied to successfully perform model-class selection (Ching et al., 2005; Ching and Chen, 2006; Muto and Beck, 2006).

1.4 Overview

This dissertation presents the application of stochastic simulation methods to problems in system identification and model-class selection in a variety of engineering problems, including several that are considered to be extremely difficult to identify. The problems presented here demonstrate the ability of stochastic simulation methods to give insight into systems that are difficult to handle with other system-identification techniques.

Chapter 2 details the Bayesian probabilistic framework that is used to perform system identification and model-class selection. Chapter 3 discusses stochastic simulation methods, introduces the specific methods to be used, and shows how these methods may be applied to implement system identification and model-class selection under the Bayesian probabilistic framework.

Chapter 4 presents an application of stochastic simulation methods to ground motion estimation using empirical attenuation relations. Chapter 5 presents a method for using stochastic simulation methods to perform structural health monitoring by updating linear structural models using incomplete modal data.

Finally, Chapter 6 deals with the identification of a class of simple yet powerful hysteretic models using recorded strong-motion data, a problem that has been demonstrated to be ill conditioned. Stochastic simulation techniques are successfully applied to identify simulated deteriorating and non-deteriorating Masing hysteretic shear-building models, and to the identification of a seven-story reinforced-concrete structure damaged during the 1994 Northridge, California, earthquake.

Chapter 2

Bayesian Model Updating and Model-Class Selection

Though the Bayesian updating approach for statistical models and model-class selection (discussed in Chapter 1) is now common in all fields of science, there are far fewer applications to system identification in engineering. Beck (1989, 1996) and Beck and Katafygiotis (1991, 1998) presented a Bayesian statistical framework for model updating and making robust predictions in the context of system identification in engineering. This chapter will illustrate how this statistical framework can be used to view system-identification problems as Bayesian updating by embedding a system model in a predictive probability model. It will also show that this concept can be generalized to apply to the problem of data-based selection between competing model classes. Additionally, some of the advantages and challenges associated with this probabilistic framework will be examined.

2.1 Model Updating

Consider the case where the output of a system, denoted $x \in \mathrm{R}^N$, is to be modeled by a model class $\mathcal{M}$that has N_p model parameters in the vector $\theta \in \Theta \in \mathrm{R}^{N_p}$, and output $y(\theta) \in \mathrm{R}^N$. Now choose some reasonable probability model for the uncertain *prediction error* $\epsilon(\theta) = x - y(\theta)$ associated with the model defined by θ. The probability model for the prediction-error determines the predictive *probability density function* (PDF) for the system output of a model defined by a given parameter vector θ. This

PDF is denoted as $p(x \mid \theta, \mathcal{M})$, which is read as "the probability density function of x, *conditioned* on the model parameter vector θ and the model class $\mathcal{M}$." In the case where the prediction error is modeled as a zero-mean, stationary, Gaussian variables with covariance matrix Σ, which is a model supported by the principle of maximum differential entropy (Jaynes, 1957; Cover and Thomas, 1991), the predictive PDF is given by

$$p(x \mid \theta, \mathcal{M}) = \frac{1}{(2\pi)^{N/2} \mid \Sigma \mid^{1/2}} \exp\left[-\frac{1}{2}\left(x - y(\theta)\right)^T \Sigma^{-1} \left(x - y(\theta)\right)\right] \tag{2.1}$$

A *Bayesian model class*, denoted $\mathcal{M}$, consists of a predictive PDF $p(x \mid \theta, \mathcal{M})$, such as the one in Equation 2.1, and a *prior* PDF $p(\theta \mid \mathcal{M})$ over the model parameters θ. The prior PDF is chosen to express the initial plausibility of the predictive PDF given by the each of the possible values of θ in Θ. This interpretation of the prior PDF is consistent with the core idea of probability logic: The probability of b, given c, is a measure of the plausibility of statement b based on the conditioning information given in statement c (Cox, 1961; Jaynes, 2003).

Now suppose a set of data $\mathcal{D}$, consisting of a set of measurements $\hat{x}$, is available. The goal of Bayesian updating is to use $\mathcal{D}$ to update the probability distribution over the parameters to give the posterior PDF $p(\theta \mid \mathcal{D}, \mathcal{M})$ based on Bayes' Theorem:

$$p(\theta \mid \mathcal{D}, \mathcal{M}) \propto p(\mathcal{D} \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{M}) \tag{2.2}$$

The constant of proportionality is $p(\mathcal{D} \mid \mathcal{M})$, called the *evidence* for model class $\mathcal{M}$, and is given by

$$p(\mathcal{D} \mid \mathcal{M}) = \int p(\mathcal{D} \mid \theta, \mathcal{M}) \; p(\theta \mid \mathcal{M}) d\theta \tag{2.3}$$

Consider the case where the prediction error of each component of x is Gaussian, uncorrelated with the other components of x, and has a variance of σ^2, that is, $\Sigma = \sigma^2 \; I_{N \times N}$. Also, every predictive PDF defined by a vector $\theta \in \Theta$ is considered equally plausible *a priori* (a uniform prior PDF on θ). In this case, Equations 2.1

and 2.2 lead to

$$p(\theta \mid \mathcal{D}, \mathcal{M}) \propto \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{\|\hat{x} - y(\theta)\|^2}{2\sigma^2}\right] \tag{2.4}$$

where $\|.\|$ denotes the Euclidean norm. It becomes apparent that finding the most probable model, i.e. the model maximizing the posterior PDF $p(\theta \mid \mathcal{D}, \mathcal{M})$, is equivalent to solving a least-squares problem. Thus, a commonly used optimization technique can easily be stated as a Bayesian model-updating problem.

Note that while the framework introduced in this section is for continuous-valued variables, it can easily be applied to discrete-valued variables by replacing the probability density functions $p(.)$ with discrete probabilities $P(.)$ and replacing the integrals with the appropriate summations.

2.1.1 System Identifiability

Beck and Katafygiotis (1998) provided rigorous definitions of *system identifiability* for a model class based on the data, $\mathcal{D}$. These definitions can be stated more informally as follows:

1. A model class $\mathcal{M}$ is *globally system identifiable* if a unique parameter vector $\hat{\theta} \in \Theta$ exists that defines the maximum likelihood model, i.e., the model maximizing the likelihood function $p(\mathcal{D} \mid \theta, \mathcal{M})$. A globally identifiable model class is shown for a two-dimensional parameter space in Figure 2.1(a).

2. A model class $\mathcal{M}$ is *locally system identifiable* if there are a finite number of parameter vectors in Θ that maximize the likelihood function, as shown in Figure 2.1(b).

3. A model class $\mathcal{M}$ is *system unidentifiable* if there is a manifold (surface) of parameter vectors in Θ that maximize the likelihood function, as shown with a one-dimensional manifold in Figure 2.1(c).

In this work, these definitions are slightly modified. Systems are classified based on the most probable model, i.e., the model maximizing the posterior PDF $p(\theta \mid \mathcal{D}, \mathcal{M})$,

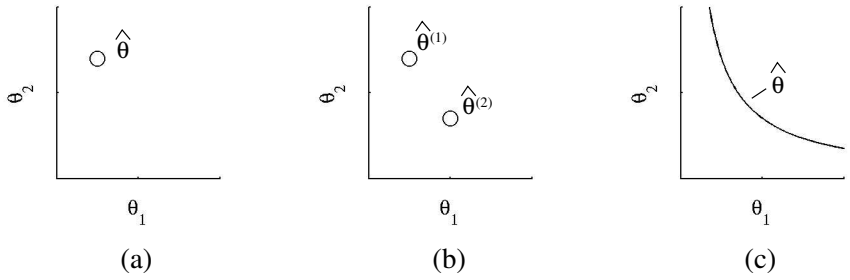


Figure 2.1: Conceptual sketch of system identifiability: (a) globally identifiable, (b) locally identifiable and, (c) unidentifiable systems

rather than the maximum likelihood model. Note that these two definitions are equivalent under the condition that the prior distribution is uniform over the parameter space. Also, for convenience, descriptions such as "globally system identifiable" will be shortened to "globally identifiable".

2.1.2 Robust Predictive PDFs

A common use of posterior probability distributions is to make inferences about future events based on past observations. Define X_1^M as a set of M potential observations of some quantity x, expressed as $X_1^M = \{x_1, \ldots, x_M\}$. Now consider the case where the first N observations of X_1^M are contained in the data set $\mathcal{D}_N$ and we wish to calculate a predictive PDF to make inferences about the remaining observations, X_{N+1}^M. However, instead of calculating the predictive PDF for a single model in $\mathcal{M}$ that is defined by the vector θ, we wish to calculate a predictive PDF that is based on the entire model class $\mathcal{M}$. The desired PDF can be calculated as follows from the theory of total probability:

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}) = \int p(X_{N+1}^M \mid \theta, \mathcal{D}_N, \mathcal{M}) p(\theta \mid \mathcal{D}_N, \mathcal{M}) d\theta \tag{2.5}$$

Note that θ alone defines the predictive PDF $p(X_{N+1}^M \mid \theta, \mathcal{M})$ of $\mathcal{M}$, so the conditioning on $\mathcal{D}_N$ in the first factor of the integrand is irrelevant and can be dropped:

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}) = \int p(X_{N+1}^M \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{D}_N, \mathcal{M}) d\theta \tag{2.6}$$

Papadimitriou et al. (2001) called the PDF defined by Equation 2.6 the *robust predictive* PDF for model class $\mathcal{M}$. The predictive PDF is robust with respect to uncertainty in θ, because this uncertainty is explictly included. If the model class $\mathcal{M}$ is globally identifiable on $\mathcal{D}_N$, then the optimal parameter vector $\hat{\theta}$ may be used to obtain an approximation of the robust predictive PDF (Beck and Katafygiotis, 1998):

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}) \approx p(X_{N+1}^M \mid \hat{\theta}, \mathcal{M}) \tag{2.7}$$

The approximation is based on Laplace's method of asymptotic approximation and it is accurate to $O\left(\frac{1}{N}\right)$, so the size of the data set N must be large for an acceptable approximation, usually the order of 50 or more.

2.1.3 Incorporating Additional Data

It is often the case that relevant data for a particular model class becomes available after an initial identification is performed. Consider the model class $\mathcal{M}$ with model parameters θ, updated with two sets of data, $\mathcal{D}_1$ and $\mathcal{D}_2$. The posterior PDF is given by

$$p(\theta \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{M}) = \frac{p(\mathcal{D}_1, \mathcal{D}_2 \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{M})}{p(\mathcal{D}_1, \mathcal{D}_2 \mid \mathcal{M})} \tag{2.8}$$

Note that the likelihood function may be expressed as

$$p(\mathcal{D}_1, \mathcal{D}_2 \mid \theta, \mathcal{M}) = p(\mathcal{D}_2 \mid \theta, \mathcal{D}_1, \mathcal{M})\ p(\mathcal{D}_1 \mid \theta, \mathcal{M}) \tag{2.9}$$

As in Equation 2.6, since θ alone defines the predictive PDF $p(x \mid \theta, \mathcal{M})$ of $\mathcal{M}$, the conditioning on $\mathcal{D}_1$ is irrelevant, such that

$$p(\mathcal{D}_1, \mathcal{D}_2 \mid \theta, \mathcal{M}) = p(\mathcal{D}_2 \mid \theta, \mathcal{M})\ p(\mathcal{D}_1 \mid \theta, \mathcal{M}) \tag{2.10}$$

Using this result with the definition of the evidence given by Equation 2.3, it can further be shown that

$$p(\mathcal{D}_1, \mathcal{D}_2 \mid \mathcal{M}) = p(\mathcal{D}_1 \mid \mathcal{M}) \int p(\mathcal{D}_2 \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{D}_1, \mathcal{M}) d\theta \tag{2.11}$$

The final result is that Equation 2.8 can be expressed as

$$\begin{aligned} p(\theta \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{M}) &= \frac{p(\mathcal{D}_2 \mid \theta, \mathcal{M})\ p(\mathcal{D}_1 \mid \theta, \mathcal{M})\ p(\theta \mid \mathcal{M})}{p(\mathcal{D}_1 \mid \mathcal{M}) \int p(\mathcal{D}_2 \mid \theta', \mathcal{M}) p(\theta' \mid \mathcal{D}_1, \mathcal{M}) d\theta'} \\ &= \frac{p(\mathcal{D}_2 \mid \theta, \mathcal{M})\ p(\theta \mid \mathcal{D}_1, \mathcal{M})}{\int p(\mathcal{D}_2 \mid \theta', \mathcal{M}) p(\theta' \mid \mathcal{D}_1, \mathcal{M}) d\theta'} \end{aligned} \tag{2.12}$$

In other words, Equation 2.12 states that updating with the entire data set is equivalent to updating with data $\mathcal{D}_1$ and using the resulting posterior PDF as the prior PDF for updating with $\mathcal{D}_2$.

This result can also be important in efficient implementation of Bayesian model updating, allowing for a reduction in the computational effort of the problem by partitioning the data, as will be demonstrated later. However, this is only possible in cases where the posterior PDF can be exactly represented.

2.2 Model-Class Selection

As discussed in Chapter 1, the concept of choosing a model class that satisfied some criteria for "model parsimony" was understood in the early 20th century (Jeffreys, 1939). However, no quantitative expression for a principle of parsimony was available until the pioneering work by Akaike (1974). Akaike recognized that maximum likelihood estimation (corresponding to pure data fitting) is insufficient for model order selection in time-series forecasting using autoregressive moving average (ARMA) models; he introduced the Akaike information criterion (AIC), which combines a measure of the data-fit, the log-likelihood of the optimal model in the model class, with a term penalizing larger numbers of uncertain (adjustable) parameters.

However, the form of the penalty term was determined in a rather ad hoc man-

ner, lacking a rigorous basis. Schwartz (1978) modified the penalty term based on a rigorous asymptotic analysis of Bayesian updating at the model-class level. This, and subsequent work (Gull, 1989; MacKay, 1992; Sivia, 1996; Beck and Yuen, 2004), demonstrated that a Bayesian approach to the problem of model-class selection automatically enforces model parsimony without the need for ad hoc penalty terms.

The goal of Bayesian model-class selection is to identify the most probable model class within a selected group of candidate model classes. It can be viewed as a generalization of the identification of the most probable values of the parameters within a model class, as discussed in Section 2.1.

Consider a pool of N_m candidate model classes, $\mathbf{M}$. Each model class $\mathcal{M}_j, j = 1 \ldots N_m$ contains N_j model parameters, denoted θ_j. The model classes are defined not only by the form for the likelihood function $p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)$, but also by the choice of prior probability distribution taken over the model parameters, $p(\theta_j \mid \mathcal{M}_j)$.

It is sometimes claimed by critics of Bayesian methods that Bayesian model-class selection assumes that the "true" model is included in the pool of candidate models, but no such assumption is actually made. Bayesian model-class selection is a framework for making rational comparisons between the models under consideration in the selected set of candidate model classes. It is the responsibility of the user to choose candidate model classes for a given problem or process. Also note that Bayesian model-class selection is not necessarily restricted to a discrete pool of candidate models; through the use of continous hyper-parameters it is possible for the problem to be continuous. An excellent example of this in a currently active area of research is the relevance vector machine (Tipping, 2000), which can be viewed as the application of Bayesian model-class selection to the support vector machine classifier (Oh and Beck, 2006).

From Bayes' theorem, we see that the probability of a particular model class $\mathcal{M}_j$ is given by

$$P(\mathcal{M}_j \mid \mathcal{D}, \mathbf{M}) = \frac{p(\mathcal{D} \mid \mathcal{M}_j, \mathbf{M}) P(\mathcal{M}_j \mid \mathbf{M})}{p(\mathcal{D} \mid \mathbf{M})} \tag{2.13}$$

where $p(\mathcal{D} \mid \mathcal{M}_j, \mathbf{M})$ is the evidence for each model class, $P(\mathcal{M}_j \mid \mathbf{M})$ is the prior

probability of model class $\mathcal{M}_j$ and $p(\mathcal{D} \mid \mathbf{M})$ is the normalizing constant given by:

$$p(\mathcal{D} \mid \mathbf{M}) = \sum_{i=1}^{N_m} p(\mathcal{D} \mid \mathcal{M}_i, \mathbf{M}) P(\mathcal{M}_i \mid \mathbf{M}) \tag{2.14}$$

It is apparent that the evidence should be the dominant factor in model-class selection, especially in the case where a non-informative prior is taken over the model classes (i.e., $P(\mathcal{M}_j \mid \mathbf{M}) = \frac{1}{N_m}$). This evidence is the normalizing constant for the posterior distribution of θ_j, as defined in Equation 2.3.

2.2.1 Information-Theoretic Approach to Model-Class Selection

Insight into how Bayesian model-class selection automatically enforces model parsimony can be obtained from considering the problem from an information-theoretic point of view, as noted in Beck and Yuen (2004) and subsequently extended by Ching et al. (2005).

Consider the natural logarithm of the evidence, $\ln[p(\mathcal{D} \mid \mathcal{M}_j)]$. Since the posterior probability distribution over θ_j is by definition normalized to unity, we can make the statement

$$\ln[p(\mathcal{D} \mid \mathcal{M}_j)] = \ln[p(\mathcal{D} \mid \mathcal{M}_j)] \int p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \tag{2.15}$$

The evidence is independent of θ_j, so it can be brought inside the integral. Making substitutions according to Bayes' Theorem (Equation 2.2):

$$\begin{aligned} \ln\left[p(\mathcal{D} \mid \mathcal{M}_j)\right] &= \int \ln\left[\frac{p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)\; p(\theta_j \mid \mathcal{M}_j)}{p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)}\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &= \int \ln[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &\quad - \int \ln\left[\frac{p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)]}{p(\theta_j \mid \mathcal{M}_j)}\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \end{aligned} \tag{2.16}$$

From this formulation, we see that the first term, which is the posterior mean of the log-likelihood, can be considered as a measure of the average data fit, while the second

term is the *relative entropy* between the prior and posterior distributions, introduced in the seminal information theory work by Shannon (1948) and later identified by Kullback and Leibler (1951) as a measure of the "distance" between probability distributions, which is why it is sometimes also termed the Kullback-Leibler divergence. The relative entropy is a measure of the information gained about the parameters θ_j from the data $\mathcal{D}$. Therefore, the evidence is comprised of a data-fit term and a term that provides a penalty against more complicated models, that is, those which extract more information from the data.

2.2.2 Evaluating the Evidence

The previous sections have shown that the key step in Bayesian model-class selection is the evaluation of the evidence for a given model class. It is possible to approximate the integral in Equation 2.3 directly with Monte Carlo simulation:

$$\int p(\mathcal{D} \mid \theta_j, \mathcal{M}_j) p(\theta_j \mid \mathcal{M}_j) d\theta_j \approx \frac{1}{N} \sum_{k=1}^{N} p(\mathcal{D} \mid \tilde{\theta}^k, \mathcal{M}_j) \tag{2.17}$$

where $\tilde{\theta}^k, k = 1 \ldots N$, are samples of θ drawn from the prior distribution $p(\theta \mid \mathcal{M}_j)$, but it is generally computationally inefficient to do so. In most cases, the region in the parameter space Θ with significant posterior probability content is very concentrated compared to the prior distribution on θ, which leads to a large variance in the estimator given by Equation 2.17 and so a very large number of samples is needed to adequately approximate the evidence.

In the cases where the model is globally identifiable and a large number N of data points is available, an asymptotic approximation to this evidence is possible. Beck and Yuen (2004) show that the application of Laplace's method of asymptotic expansion to the integrals in Equation 2.16 yields the following approximations, when

N is large:

$$\int \ln[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \approx \ln[p(\mathcal{D} \mid \hat{\theta}_j, \mathcal{M}_j)] \tag{2.18}$$

$$\int \ln\left[\frac{p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)]}{p(\theta_j \mid \mathcal{M}_j)}\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \approx \ln\left[(2\pi)^{\frac{N_j}{2}} \left|\mathbf{H}(\hat{\theta}_j)\right|^{-1}\right] - \ln[p(\hat{\theta}_j \mid \mathcal{M}_j)] \tag{2.19}$$

where $\hat{\theta}_j$ is the vector of the most probable values of θ_j and $\mathbf{H}(\hat{\theta}_j)$ is the Hessian matrix of $-\ln[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j) p(\theta \mid \mathcal{M})]$ evaluated at $\hat{\theta}_j$. The approximation used here is similar to the one used to derive the approximation in Equation 2.7, which implies that the posterior PDF on θ is well approximated by a Gaussian distribution:

$$p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) \approx \frac{1}{(2\pi)^{\frac{N_j}{2}} \left|\mathbf{H}(\hat{\theta}_j)\right|^{-\frac{1}{2}}} \exp\left[-\frac{1}{2}(\theta_j - \hat{\theta}_j)^T \mathbf{H}(\hat{\theta}_j)(\theta_j - \hat{\theta}_j)\right] \tag{2.20}$$

for θ_j in a neighborhood of $\hat{\theta}_j$ of size $O(N^{-\frac{1}{2}})$.

The quantity in Equation 2.19, which is the asymptotic approximation of the relative entropy, was identified by Gull (1989) as the logarithm of the "Ockham factor," a penalty against parameterization, denoted here by K_j. Thus, the log evidence for $\mathcal{M}_j$ is approximately given by

$$\ln[p(\mathcal{D} \mid \mathcal{M}_j)] = \ln[p(\mathcal{D} \mid \hat{\theta}_j, \mathcal{M}_j)] - K_j \tag{2.21}$$

Beck and Yuen (2004) demonstrated the effectiveness of this asymptotic approach for globally identifiable model classes for dynamical systems.

2.2.3 Bayesian Model Averaging

So far, it has been assumed that the goal of model-class selection is to select a single model class for use in calculating the robust predictive PDF for a system. This is not, however, required. If desired, the robust predictive PDF can be expanded to include every model class predictive PDF over the entire candidate model pool. For example,

the predictive PDF for some output Y given the input X, and the previously available data $\mathcal{D}$ over the pool of N_m candidate models $\mathbf{M}$, is given by

$$p(Y \mid X, \mathcal{D}, \mathbf{M}) = \sum_{j=1}^{N_m} P(\mathcal{M}_j \mid \mathcal{D}, \mathbf{M}) \int p(Y \mid X, \theta_j, \mathcal{M}_j)\ p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \quad (2.22)$$

This is essentially weighting the predictive PDF for each model class $\mathcal{M}_j$, $j = 1, \ldots, N_m$, with the posterior model class probabilities, $P(\mathcal{M}_j \mid \mathcal{D}, \mathbf{M})$. In practice, it is often the case that one model class dominates; however, Bayesian model averaging may be useful in cases where there are multiple model classes with significant probabilites and there are substantial differences in the predictions (such as different predicted structural behavior at unobserved degrees of freedom).

2.2.4 Influence of Prior PDFs

Prior PDFs generally do not significantly influence the form of the posterior PDF if there is a sufficient amount of data and the model class is globally identifiable, except, of course, in cases where regions of high probability content are completely excluded (i.e., the value of the prior PDF is zero for those regions). However, if we consider the information extraction term in Equation 2.16, it becomes apparent that the chosen prior PDF may have a significant influence on model-class selection results.

This situation does not seem unreasonable when the chosen prior PDF is developed from previous work, experience, etc. However, it is often the case that the user will have no information about a particular parameter or set of parameters. In this case, it is desirable to select a prior PDF that accurately expresses this state, sometimes referred to as an "ignorance prior," or more appropriately described as a "non-informative" prior; that is, a prior distribution that does not bias the posterior in any way.

Naively, it would seem that the best choice for a prior PDF would be a uniform distribution over a very large interval, or some similarly broad distribution. However, while such a choice of prior PDF may not bias the identification within a given model

class, it may create a bias in the model-class selection process; when widely dispersed priors are used, Bayesian model-class selection tends to favor simpler model classes, regardless of the quality of the data fit.

This behavior can be explained using the information-theory interpretation of model-class selection. When widely dispersed priors are used, the information-gain term dominates. Equation 2.16 may be re-written as

$$\begin{aligned} \ln\left[p(\mathcal{D} \mid \mathcal{M}_j)\right] &= \int \ln\left[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &\quad - \int \ln\left[p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &\quad + \int \ln\left[p(\theta_j \mid \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \end{aligned} \tag{2.23}$$

Now consider a uniform prior expressed as $p(\theta_j \mid \mathcal{M}_j) = V_j^{-1}$, where V_j represents the volume in the parameter space Θ_j (assumed to be bounded). Substituting this into the previous equation,

$$\begin{aligned} \ln\left[p(\mathcal{D} \mid \mathcal{M}_j)\right] &= \int \ln\left[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &\quad - \int \ln\left[p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j - \ln V_j \end{aligned} \tag{2.24}$$

Note that for very widely dispersed priors, where V_j grows with the dimension of the parameter space, the evidence will favor less complicated models (fewer unknown parameters) regardless of the data. From an information-theory viewpoint, the issue is that for widely dispersed priors PDFs, there is a huge gain of information (an infinite amount of information gained for improper priors, where V_j is infinite), which makes the contribution to the evidence of the average data-fit term negligible. Thus, widely dispersed priors are not non-informative in the model-class selection sense, as they create a bias toward simpler models. A well-chosen non-informative prior would therefore be a prior PDF that does not bias the evidence.

Jeffreys (1932, 1939) proposed a non-informative prior PDF based on considerations of invariance under re-parameterization (e.g., specifying the prior PDF for θ is

equivalent to specifying the prior for θ^2), which has the following form:

$$p(\theta \mid \mathcal{M}) = \frac{\mid I(\theta \mid \mathcal{M}) \mid^{\frac{1}{2}}}{\int_{\Theta} \mid I(\theta' \mid \mathcal{M}) \mid^{\frac{1}{2}} d\theta'} \tag{2.25}$$

where $I(\theta \mid \mathcal{M})$ is the *Fisher Information Matrix* (Fisher, 1925), defined as:

$$[I(\theta \mid \mathcal{M})]_{ij} = \int p(x \mid \theta, \mathcal{M}) \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln[-p(x \mid \theta, \mathcal{M})] dx \tag{2.26}$$

This is known as the *Jeffreys prior*. It is widely accepted as a reasonable non-informative prior PDF in the single-parameter case; however, implementation in multivariate problems is difficult and there is no clear consensus on an appropriate choice in such cases (Bernardo and Smith, 2000). Jaynes (1968, 2003) presents arguments for non-informative prior PDFs based on maximum entropy considerations, but again there is no widespread agreement on what is still a very active research topic in Bayesian probability theory (see Bernardo and Smith (2000) for a summary).

Consequently, non-informative prior PDFs are not used in this work. Prior PDFs are assigned to parameters based on assumptions that are considered to reflect reasonable engineering judgement. Since these assumptions are explicitly stated, they can easily be examined, criticized, and, if necessary, revised.

2.2.5 Comparing Bayesian Model Selection to Other Model-Class Selection Criteria

The approximation to the evidence given by Equation 2.21 can be used to make comparisons between Bayesian model-class selection and other well-known model-class selection criteria. If we assume a Gaussian prior distribution $p(\theta \mid \mathcal{M}) = N(\mu_0, \Sigma_0)$ and a globally identifiable model class based on a large amount of data, then the relative entropy approximation K is given by

$$K = \frac{1}{2}\left(\ln\left[\frac{\det \Sigma_0}{\det \Sigma}\right] + Tr[\Sigma_0^{-1}\Sigma - I_{N_p \times N_p}] + (\mu - \mu_0)^T \Sigma_0^{-1}(\mu - \mu_0)\right) \tag{2.27}$$

where $\Sigma = \mathbf{H}(\hat{\theta}_j)^{-1}$ and $\mu = \hat{\theta}_j$. If it is further assumed that the user-selected covariance matrix for the prior PDF, Σ_0, is chosen to be diagonal and that $(\sigma_i^0)^2$ and $(\sigma_i)^2$, $i = 1, \ldots, N_p$ are used to denote the principal variances of Σ_0 and Σ, respectively, then

$$K = \frac{1}{2} \sum_{i=1}^{N_p} \left(\ln \left[\frac{(\sigma_i^0)^2}{(\sigma_i)^2} \right] + \frac{(\sigma_i)^2}{(\sigma_i^0)^2} - 1 + \frac{(\mu_i - \mu_i^0)^2}{(\sigma_i^0)^2} \right) \tag{2.28}$$

The variance of the posterior PDF decreases as the amount of the data increases (i.e., there is less uncertainty in the estimate as more data is used for updating), such that $(\sigma_i)^2 = O(1/N)$ (Beck and Katafygiotis, 1998). Therefore, the principal variances of the posterior PDF can be expressed as

$$(\sigma_i)^2 = \frac{k_i^2}{N} \tag{2.29}$$

Substituting this expression into Equation 2.27, we find that

$$\begin{aligned} K &= \frac{1}{2} \sum_{i=1}^{N_p} \left(\ln \left[N \frac{(\sigma_i^0)^2}{k_i^2} \right] + \frac{1}{N} \frac{k_i^2}{(\sigma_i^0)^2} - 1 + \frac{(\mu_i - \mu_i^0)^2}{(\sigma_i^0)^2} \right) \\ &= \frac{1}{2} N_p \ln N + \frac{1}{2} \sum_{i=1}^{N_p} \left(2 \ln \left[\frac{\sigma_i^0}{k_i} \right] - 1 + \frac{(\mu_i - \mu_i^0)^2}{(\sigma_i^0)^2} \right) + \frac{1}{N} \sum_{i=1}^{N_p} \frac{k_i^2}{(\sigma_i^0)^2} \\ &= \frac{1}{2} N_p \ln N + O(1) \end{aligned} \tag{2.30}$$

As $N \to \infty$, the $O(1)$ terms may neglected, so that the approximation of the log-evidence is

$$\ln[p(\mathcal{D} \mid \mathcal{M}_j)] \approx \ln[p(\mathcal{D} \mid \hat{\theta}_j, \mathcal{M}_j)] - \frac{1}{2} N_p \ln N \tag{2.31}$$

which is the form of the Bayesian information criterion (BIC) developed by Schwartz (1978). Thus, the BIC is an asymptotic estimator of the log-evidence.

Now compare the approximation to the log-evidence given in Equation 2.21 to Akaike's Information Criterion (AIC) (Akaike, 1974), which is given in a generalized

form by

$$AIC(\hat{\theta}_j, \mathcal{M}_j) = \ln[p(\mathcal{D} \mid \hat{\theta}_j, \mathcal{M}_j)] - cN_p \tag{2.32}$$

we see from Equation 2.28 that to achieve a penalty term in AIC would require that $(\sigma_i)^2 \approx (\sigma_i^0)^2$, in order to get $K = cN_p$ where $c = O(1)$ with respect to the data size N, but this contradicts the fact that the variance of the posterior PDF depends on the amount of data used for updating. Therefore, the AIC is not consistent with the log-evidence and, given that it implies behavior that is known to be inconsistent with posterior PDFs for globally identifiable systems, is a poor choice for a selection criteria.

Another model selection criterion is known as the minimum description length (MDL) principle, developed by Rissanen (1978, 2005) in the field of information theory. In this approach, a probability distribution is viewed as a coding for an object, such that the "length" of the code is identified with the negative binary (base two) logarithm of the probability of an object. The goal of the MDL approach is to find the shortest code length that represents the available data. Rissanen (1978) showed that as the number of data points N approaches infinity, the MDL principle gives identical results to the BIC. Grünwald (2005), gives an even more powerful result for an implementation of the MDL principle known as the normalized maximum likelihood (NML) criterion, showing that, when Bayesian model-class selection is performed using a Jeffreys prior (see Equations 2.25 and 2.26) over the model parameters, the two methods are *indistinguishable*.

2.3 Example

To illustrate some of the concepts discussed in this chapter, a simple example is presented. Consider a system where output $x(t) = \cos\left(\frac{\pi}{2}t\right)$ on the interval $t \in [0, 1]$. The available data $\mathcal{D}$ is a set of $N = 100$ measurements of x at times $\hat{t}$ drawn from a uniform distribution between 0 and 1. Gaussian white noise with a standard deviation of 0.1 is added to the observations to form the vector $\hat{x}$, which is shown

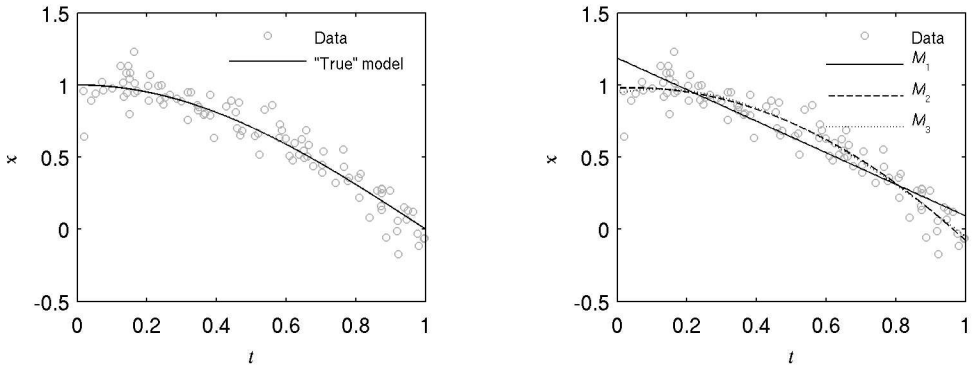


Figure 2.2: Data for example problem plotted with (a) the output of the "true" system and (b) the predicted output for the maximum likelihood model in each model class.

in Figure 2.2(a). The candidate model classes will be first-, second- and, third-order polynomials denoted $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$, respectively. Note that the"true" model class is not contained in the candidate model pool. The prediction-error variance for each measurement is assumed to be uncorrelated with a common variance σ, so that the likelihood function is of the form shown in Equation 2.4, where the model output at time t is given by

$$y(t \mid \theta) = \sum_{i=1}^{N_p} \theta_{i-1} t^{i-1} \tag{2.33}$$

where N_p is the number of terms in the polynomial and the polynomial cofficients θ_i are uncertain parameters to be updated, along with σ.

The prior PDF for each coefficient in each model class is taken to be a zero-mean Gaussian with standard deviation σ_0. The prior PDF for the prediction-error variance is uniformly distributed between 0 and 1 for all model classes.

The optimal model in each model class is determined by numerical optimization of the log of the product of the likelihood function and the prior PDF, $p(\mathcal{D} \mid \theta, \mathcal{M})p(\theta \mid \mathcal{M})$, which is proportional to the posterior PDF. Standard deviations are estimated by approximating the covariance matrix of the posterior PDF as the inverse of the Hessian matrix of the negative log of the posterior PDF, as in Equation 2.20. Table 2.1 shows the identified optimal models in each model class as the standard deviation of the prior PDF σ_0 is increased, as well as the maximum likelihood estimate for each

Model Class	σ_0	θ_3	θ_2	θ_1	θ_0	σ
$\mathcal{M}_1$	0.1			-0.6310	0.9174	0.1899
				(0.0574)	(0.0347)	(0.0148)
	1			-1.0896	1.1836	0.1292
				(0.0441)	(0.0259)	(0.0091)
	10			-1.0929	1.1855	0.1292
				(0.0441)	(0.0259)	(0.0091)
	100			-1.0929	1.1855	0.1292
				(0.0441)	(0.0259)	(0.0091)
	MLE			-1.0929	1.1855	0.1292
$\mathcal{M}_2$	0.1		0.6747	-0.2848	0.9946	0.1105
			(0.0657)	(0.0678)	(0.0221)	(0.0081)
	1		-1.1295	0.0719	0.9820	0.0990
			(0.1346)	(0.1426)	(0.0317)	(0.0070)
	10		-1.1487	0.0911	0.9789	0.0990
			(0.1372)	(0.1454)	(0.0317)	(0.0070)
	100		-1.1488	0.0913	0.9789	0.0990
			(0.1372)	(0.1454)	(0.0317)	(0.0070)
	MLE		-1.1488	0.0913	0.9789	0.0990
$\mathcal{M}_3$	0.1	-0.4209	-0.4097	-0.1878	0.9667	0.1040
		(0.0723)	(0.0802)	(0.0688)	(0.0219)	(0.0075)
	1	0.0332	-1.1795	0.0921	0.9802	0.0.990
		(0.3761)	(0.5826)	(0.2700)	(0.0373)	(0.0070)
	10	0.5754	-2.0388	0.4659	0.9433	0.0985
		(0.5401)	(0.8467)	(0.3804)	(0.0459)	(0.0070)
	100	0.5820	-2.0572	0.4740	0.9424	0.0985
		(0.5431)	(0.8515)	(0.3824)	(0.00461)	(0.0070)
	MLE	0.5871	-2.0574	0.4739	0.9424	0.0985

Table 2.1: Most probable parameter estimates for each model class. Standard deviations are shown in parentheses. "MLE" indicates the Maximum Likelihood estimate for the parameters (no influence of the prior PDF).

model class (denoted "MLE"), which is not influenced by the prior PDF.

The results show that the optimal model is only significantly influenced by the choice of prior PDF when values of σ_0 are small, which would tend to restrict the parameter values to the neighborhood around the origin. The difference between the most probable parameter estimates and the maximum likelihood estimates are small for larger values of σ_0, compared to the estimated standard deviations for the posterior PDF. Figure 2.3(a) shows the prior PDF, and Figures 2.3(b)–(d) show the posterior PDF for each model class when $\sigma_0 = 10$, demonstrating that the more complicated model classes have greater uncertainty associated with the parameter estimates, since

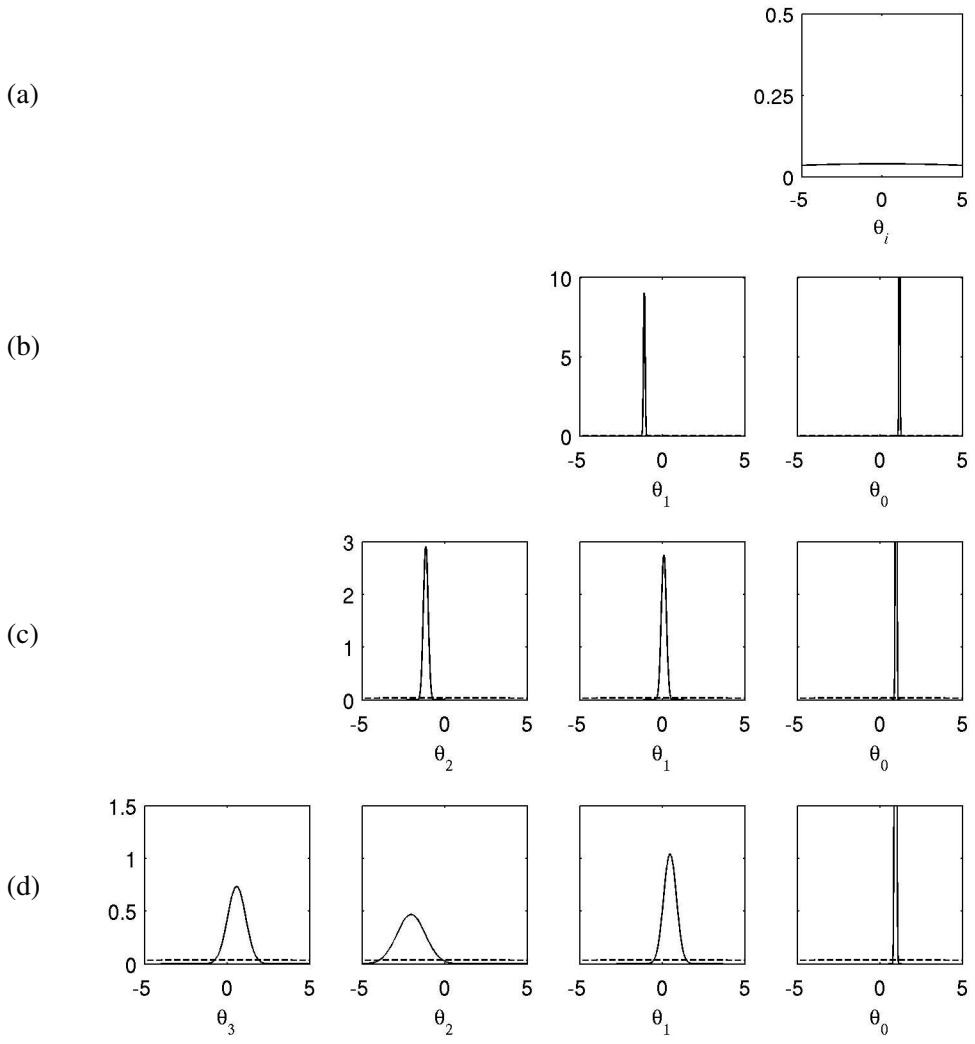


Figure 2.3: (a) Prior PDF for all parameters and posterior PDFs for the polynomial coefficients of model classes (b) $\mathcal{M}_1$, (c) $\mathcal{M}_2$, and (d) $\mathcal{M}_3$, for the case where the standard deviation of the prior PDF for each parameter is $\sigma_0 = 10$.

the same data is being used to identify more parameters. Also note that the constant term θ_0 is well defined for all model classes, which is to be expected since none of the other terms contribute significantly to the response for small t, thus most of the information in this region is used to pin down θ_0.

Model class $\mathcal{M}_3$ has the best fit to the data (the most probable value for σ corresponds to the root-mean-square value of the error between data and the predicted model ouput), though it is only slightly better than $\mathcal{M}_2$. This is illustrated by Figure 2.2(b), which shows the predicted output for the maximum likelihood model

σ_0	Model Class	$\ln[p(\mathcal{D} \mid \hat{\theta}_j, \mathcal{M}_j)]$	K_j	$\ln[p(\mathcal{D} \mid \mathcal{M}_j)]$	Probability
10	$\mathcal{M}_1$	62.78	15.87	46.91	0.0000
	$\mathcal{M}_2$	89.34	20.95	68.39	0.9153
	$\mathcal{M}_2$	89.92	23.91	66.01	0.0847
100	$\mathcal{M}_1$	62.78	20.46	42.32	0.0000
	$\mathcal{M}_2$	89.34	27.85	61.49	0.9906
	$\mathcal{M}_2$	89.92	33.08	56.83	0.0094
10000	$\mathcal{M}_1$	62.78	29.67	33.11	0.0000
	$\mathcal{M}_2$	89.34	41.66	47.68	0.9999
	$\mathcal{M}_2$	89.92	51.51	38.41	0.0001
1000000	$\mathcal{M}_1$	62.78	38.88	23.90	0.0000
	$\mathcal{M}_2$	89.34	55.48	33.86	1.0000
	$\mathcal{M}_3$	89.92	69.93	19.99	0.0000

Table 2.2: Model-class selection results for the example problem.

in each model class.

The evidence is estimated using the asymptotic approximation given by Equations 2.18, 2.19, and 2.21. Results are shown in Table 2.2 for several values of σ_0. We see that model class $\mathcal{M}_2$ dominates. Model class $\mathcal{M}_3$ has a small non-zero probabilty, as it benefits somewhat from a better data-fit term. However, as σ_0 (and therefore the information gain) increases, the more complicated model becomes less probable. Since the data-fit term is essentially constant, while the information gain term grows with σ_0, model class $\mathcal{M}_1$ will be preferred for a very widely dispersed prior PDF (as discussed in Section 2.2.4). However, these results demonstrate that even an extremely large prior such as $\sigma_0 = 1 \times 10^6$ still results in a preference for model class $\mathcal{M}_2$.

2.4 Conclusions

Bayesian inference provides a rational, transparent framework, consistent with the axioms of probability, for combining data with prior knowledge to update models, make predictions that are robust to uncertainty, and choose between competing model classes. The inclusion of the prior PDF on the parameters in defining a model class, rather than as a separate element of the identification process, is a new concept, and important due to the effect that prior PDFs can have on model-class selection. An

information-theoretic interpretation of model-class selection aids in understanding this effect, and also explains the Lindley paradox in Bayesian model-class selection. This interpretation will play an important role in understanding several of the results for model-class selection presented in later chapters. A simple example problem is presented to demonstrate these concepts.

Chapter 3

Stochastic Simulation Methods

The Bayesian methods for model updating and model-class selection presented in Chapter 2 are often difficult to implement for ill-conditioned problems, which often arise in system identification. The asymptotic approach for updating and model-class selection is most useful when there is a large amount of data and the model class is globally identifiable, although it can be applied in locally identifiable cases and even unidentifiable cases where the manifold of optimal models is of low dimension (Papadimitriou et al., 2001; Katafygiotis and Lam, 2002). However, it requires a possibly non-convex high-dimensional optimization to determine the optimal parameter vectors, which can be computationally challenging. Additionally, the asymptotic approximation for the evidence requires evaluating the Hessian matrix of the log-likelihood function, which may be difficult to perform analytically and computationally expensive to accurately approximate numerically.

Given these challenges, in recent years attention has been focused on stochastic simulation methods, especially Markov Chain Monte Carlo methods, for Bayesian updating, prediction, and model-class selection.

3.1 Overview of Stochastic Simulation Methods

The goal of stochastic simulation methods is to generate samples that are distributed according to some target PDF. In the specific case of Bayesian model updating, the target is the posterior PDF, $p(\theta \mid \mathcal{D}, \mathcal{M})$.

While a variety of simulation methods is available, many of the methods are not useful for Bayesian updating. In this work, we focus specifically on Markov Chain Monte Carlo (MCMC) methods. Neal (1993), Gilks et al. (1996), Robert and Casella (1999), and MacKay (2003) provide more comprehensive overviews of this topic. The samples generated by MCMC methods are from a Markov chain whose stationary distribution is the target PDF. The great advantage of these methods is that non-normalized PDFs can be sampled, so that the samples may be drawn from the posterior PDF without evaluating the evidence.

Note that efficient sampling methods are available for most well-known probability distributions (the uniform distribution, the Gaussian distribution, the gamma distribution, etc.) and are implemented in commonly used software packages, such as MATLAB. Thus, when sampling from a distribution with a known form is called for, in this chapter or the following chapters, it may be safely assumed to be easily implemented.

3.1.1 Metropolis-Hastings Sampler

One of the best-known and most commonly implemented MCMC methods is the Metropolis-Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970). The algorithm is presented in detail in Appendix A.1.

Denote the ith sample generated by the Markov chain by $\tilde{\theta}_i$. To obtain $\tilde{\theta}_{i+1}$, a proposal PDF $q(\theta^* \mid \tilde{\theta}_i)$ is used to generate a candidate sample. The sample is accepted with probability $\min(1, r)$ where the acceptance ratio r is given by

$$\begin{aligned} r &= \frac{p(\theta^* \mid \mathcal{D}, \mathcal{M})\ q(\tilde{\theta}_i \mid \theta^*)}{p(\tilde{\theta}_i \mid \mathcal{D}, \mathcal{M})\ q(\theta^* \mid \tilde{\theta}_i)} \\ &= \frac{p(\mathcal{D} \mid \theta^*, \mathcal{M})\ p(\theta^* \mid \mathcal{M})\ q(\tilde{\theta}_i \mid \theta^*)}{p(\mathcal{D} \mid \hat{\theta}_i, \mathcal{M})\ p(\tilde{\theta}_i \mid \mathcal{M})\ q(\theta^* \mid \tilde{\theta}_i)} \end{aligned} \tag{3.1}$$

If the candidate sample is not accepted, then $\tilde{\theta}_{i+1} = \tilde{\theta}_i$. This procedure is repeated until the desired number of samples has been generated.

The effectiveness of the M-H algorithm is heavily dependent on the proposal PDF.

If there is not good insight into a good choice for the proposal PDF, the algorithm may perform poorly, particularly in higher dimensions.

3.1.2 Gibbs Sampler

Another commonly implemented method is the Gibbs sampler, which is actually a special case of the M-H algorithm but without the need to choose a proposal PDF (Geman and Geman, 1984; Gelfand et al., 1990). The algorithm is presented in detail in Appendix A.2. A sequence of draws are made from a series of conditional PDFs, one for each component (or group of components) of the parameter vector θ. For example, if the vector of model parameters θ is divided into two sub-groups, θ_1 and θ_2, an initial value $\tilde{\theta}_2^{(0)}$ is selected for θ_2, and the sample $\tilde{\theta}_1^{(1)}$ is drawn from the conditional distribution $p(\theta_1 \mid \tilde{\theta}_2^{(0)}, \mathcal{D}, \mathcal{M})$, which in turn is used to obtain $\tilde{\theta}_2^{(1)}$, which is sampled from $p(\theta_2 \mid \tilde{\theta}_1^{(1)}, \mathcal{D}, \mathcal{M})$. The procedure is repeated until the desired number of samples is obtained.

It can be shown that the updating step is an M-H step that is always accepted (that is, an acceptance probability of unity). In general, the rate of convergence may be slower than for other types of updates, but in cases where the full posterior PDF can be decomposed into conditional PDFs that are easily sampled, the algorithm can be extremely efficient. Additionally, the effective dimension of the problem is reduced to the number of sub-groups, allowing for the treatment of model classes with a large number of uncertain parameters.

3.1.3 Transitional Markov Chain Monte Carlo

A remaining challenge associated with model updating by stochastic simulation is the fact that, unless the data is sparse, the posterior PDF occupies a much smaller volume in the parameter space than the prior PDF over the parameters. Consequently, for higher-dimensional parameter spaces, it may still be very difficult to draw samples that cover all the regions of high-probability content. For this reason, Beck and Au (2002) proposed a method where the model is gradually updated by using the

M-H algorithm to sample from a sequence of target PDFs, where each target PDF is the posterior PDF based on an increasing fraction of the available data. In this manner, the target PDF gradually converges from the broad prior PDF to the concentrated posterior PDF. The samples from each intermediate PDF are used to form a kernel density function (Silverman, 1986), an estimator for a density function that is composed of a weighted sum of Gaussian distribution centered at samples in the parameter space from the previous PDF in the sequence. The kernel density function is used as a global proposal PDF for applying the M-H algorithm to the next level of sampling.

Ching and Chen (2006) modified this approach to develop what they call they transitional Markov chain Monte Carlo (TMCMC) method. This technique also uses a series of intermediate PDFs. However, rather than updating with part of the available data, the entire data set is used but its full effect is diluted by taking the target PDF for the mth level of the sampler to be proportional to $p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_m} p(\theta \mid \mathcal{M})$, where $0 \leq \beta_m \leq 1$; here, $\beta_0 = 0$ gives the initial target distribution proportional to the prior PDF and $\beta_M = 1$ for the final level of the sampler gives a target distribution proportional to the posterior PDF. Figure 3.1 shows how the histograms for a single parameter converge during a run of the TMCMC algorithm for increasing values of β_m, which can be thought of as the percentage of the full effect of the data applied at a given level of the sampler. Due to the conceptual similarities between this approach and the simulated annealing approach (Neal, 1993; Fishman, 1996), β_m will be referred to as the tempering parameter. In TMCMC, re-sampling is used between levels to improve the rate of convergence. The TMCMC algorithm is presented in detail in Appendix A.3. Another difference between the TMCMC algorithm and the approach of Beck and Au (2002) is in the application of the M-H algorithm. Rather than using a global proposal PDF based on a kernel density constructed from the samples from the previous level, a local proposal PDF is used in what is essentially a random walk from the previous samples in the parameter space.

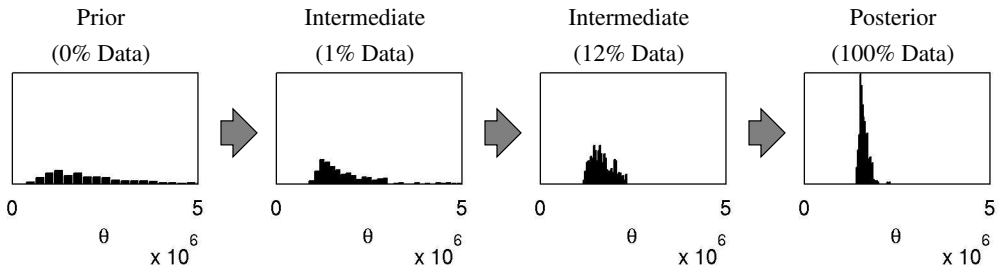


Figure 3.1: Convergence of the TMCMC algorithm for a single parameter

3.1.4 Determining Chain Length

The starting points for Markov chains used for Bayesian model updating are typically drawn from the prior PDFs, and therefore it is very likely that these starting points will be in a low-probability region of the posterior PDF. This region would then be disproportionately represented in the resulting set of samples unless the number of generated samples is very large. It is a common practice among MCMC users to discard some number of samples generated at the start of the simulation, referred to as the *burn-in period* for a Markov chain. The user is, in effect, waiting until it is believed that the samples are being drawn from the stationary distribution (the target PDF) before "recording" the generated samples. There is no widely accepted criterion for setting the length of burn-in periods. For the purposes of this work, the Markov chains for a given problem are determined by visual inspection. Figure 3.2(a) shows samples generated by stochastic simulation methods for a model parameter. It should be apparent that the burn-in period is set such that the effects of the initial conditions of the sampler have "died out".

The other issue in determining chain length is determining how many samples are necessary to accurately represent the qunatities of interest, such as the mean and standard deviation of the target PDF. This issue is sometimes referred to as the *convergence* of a Markov Chain. There have been some suggested diagnostic criteria to determine whether a Markov chain has converged, such as the Gelman-Rubin statistic (Gelman and Rubin, 1992), which compares the variance within parallel Markov chains to the variance between the chains, but there is no generally accepted standard. In this work, convergence is determined by inspection of quantities of

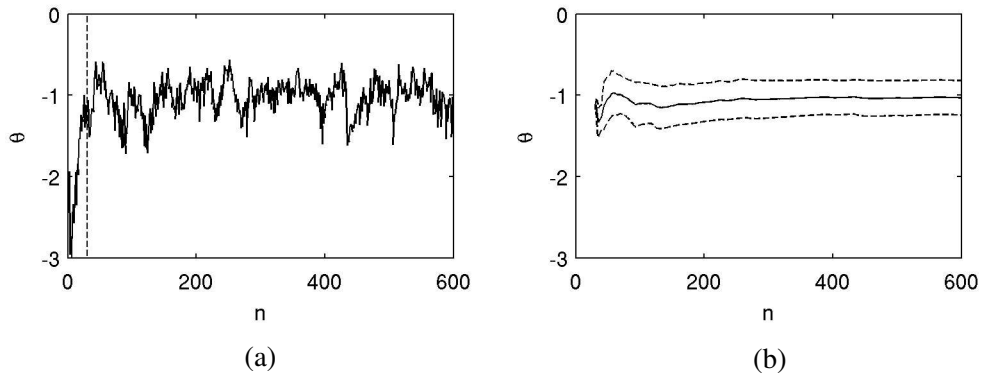


Figure 3.2: (a) Samples generated by stochastic simulation. The index n denotes the step number of the Markov chain. The dashed line indicates the end of the burn-in period at $n = 30$. (b) The evolution of the mean and standard deviation as more samples are added. The solid line indicates the mean, and the dashed lines indicate the mean plus and minus the standard deviation.

interest. For example, Figure 3.2(b) shows the evolution of the mean and standard deviation estimated from the Markov chain in Figure 3.2(a) as more samples are added. The samples shown appear to have converged to the mean and standard deviation of the target PDF after approximately 400 samples.

3.2 Representing Posterior PDFs with Samples

Once a set of samples has been generated for a posterior PDF, can the vector of most probable parameters (the vector $\hat{\theta}$ maximizing the posterior PDF) be estimated from the samples? One simple approach might be to take the sample mean for each parameter, or, if the distribution is non-symmetric, fitting a curve to a histogram of the samples of each parameter. However, this method is not always satisfactory, because the posterior PDF is a *joint PDF* over the parameters. That is, if the model parameter vector θ for model class $\mathcal{M}$ consists of components θ_1 and θ_2, the posterior PDF $p(\theta \mid \mathcal{DM})$ is the joint PDF $p(\theta_1, \theta_2 \mid \mathcal{DM})$. However, if the samples are projected onto a 1-D space (i.e., taking a histogram), the resulting density function is the *marginal PDF* for the parameter in question, integrating out the effect of the other parameters. In the two-parameter case, the marginal posterior PDF for θ_1 is

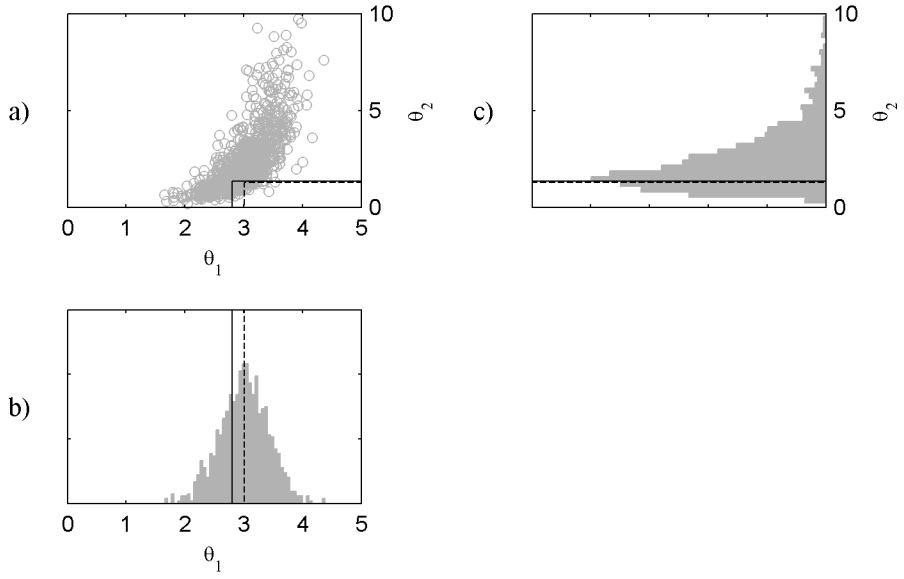


Figure 3.3: (a) samples generated for a system, and histograms of the samples for (b) θ_1, and (c) θ_2, which approximate the marginal PDFs. The solid line represents the parameter values maximizing the joint PDF, while the dashed lines represent parameter values maximizing each of the marginal PDFs.

given by

$$p(\theta_1 \mid \mathcal{D}, \mathcal{M}) = \int p(\theta_1, \theta_2 \mid \mathcal{D}, \mathcal{M}) d\theta_2 \tag{3.2}$$

If the joint posterior PDF is well approximated by a multi-dimensional Gaussian distribution, or some other symmetric distribution, the vector of most probable parameters (maximizing the joint PDF) may be determined by identifying the most probable values for each component separately (maximizing each of the marginal PDFs). However, if this is not the case, the results can be misleading.

Figure 3.3(a) shows samples generated from a PDF where one of the parameters is Gaussian-distributed while the other is lognormally distributed, with significant correlation between the parameters. Figures 3.3(b) and (c) show the histograms of the samples for θ_1 and θ_2, respectively. The parameter values maximizing each of the marginal PDFs (by curve-fitting of the histograms) are indicated by the dashed lines. However, these values differ from the set of parameter values maximizing the joint PDF, indicated by the solid lines.

In this case, determining the most probable parameters from samples would involve fitting some type of multi-dimensional function to the samples. However, if such a representation is accurate and easily obtainable, there would be little need to apply stochastic simulation methods. More sophisticated methods for estimating the posterior PDF from samples, such as the kernel density estimator (discussed briefly in Section 3.1.3) are available, but may be difficult to implement in higher-dimensional parameter spaces. In this work, it is judged that cases in which the posterior PDF cannot be accurately and efficiently represented by a single vector of parameter values are where identification with stochastic simulation is most useful. In this case, the posterior PDF is best represented by the entire set of samples rather than by summary statistics. Therefore, whenever most probable parameter values are estimated from samples in this work, it is done by maximizing the marginal PDF of each parameter. Though this will inevitably result in discrepancies, there is still some value in comparisons to most probable parameter values obtained through optimization of the joint posterior PDF for parameters that are globally identifiable.

3.3 Model-Class Selection Using Stochastic Simulation

One of the great advantages of the simulation methods discussed in Section 3.1 is the ability to sample posterior distributions without calculating the normalizing constant. However, the model-class selection requires the calculation of this constant, which is given by the evidence for the model class based on the available data.

The essential problem in evaluating the evidence using stochastic simulation methods is that simulation methods do not provide the form of the target PDF, but rather samples drawn from the target distribution. However, Ching et al. (2005) introduced the following method for estimating the evidence from stochastic simulation results.

Equation 2.16 can be re-written as follows:

$$\begin{aligned} \ln p(\mathcal{D} \mid \mathcal{M}_j) &= \int \ln \left[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)\ p(\theta_j \mid \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &\quad - \int \ln p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \\ &= \int \ln p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)\ p(\theta_j \mid \mathcal{M}_j) p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j + H[p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)] \end{aligned} \tag{3.3}$$

The first term can be approximated using N samples drawn from the target distribution, $\tilde{\theta}_j^k, k = 1 \ldots N$:

$$\int \ln \left[p(\mathcal{D} \mid \theta_j, \mathcal{M}_j)\ p(\theta_j \mid \mathcal{M}_j)\right] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \approx \frac{1}{N} \sum_{k=1}^{N} ln[p(\mathcal{D} \mid \hat{\theta}_j^k, \mathcal{M}_j)\ p(\tilde{\theta}_j^k \mid \mathcal{M}_j)] \tag{3.4}$$

The second term in Equation 3.3 is the Shannon information entropy (Shannon, 1948) for the posterior distribution:

$$H[p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)] \equiv - \int \ln p(\theta_j \mid \mathcal{D}, \mathcal{M}_j)] p(\theta_j \mid \mathcal{D}, \mathcal{M}_j) d\theta_j \tag{3.5}$$

There are several algorithms available for estimating the information enropy of a distribution using samples, such as the nearest-neighbor approach (Kozachenko and Leonenko, 1987). However, this approach, which has been shown to work with samples generated using the GS algorithm (Ching et al., 2005), cannot be used with methods that generate repeated samples, such as the M-H algorithm.

The algorithms for model updating presented in Section 3.1.3 can also be used to estimate the evidence for a model class. In the TMCMC algorithm, when generating samples for the mth target PDF, re-sampling is performed on the samples from the $(m-1)$th level, $\tilde{\theta_k}^{(m-1)}, k = 1, \ldots, N$. The re-sampling weight for each sample, $w\left(\tilde{\theta_k}^{(m-1)}\right)$, is the ratio of the target PDFs for the mth and $(m-1)$th levels, evaluated

at $\tilde{\theta_k}^{(m-1)}$:

$$\begin{aligned} w\left(\tilde{\theta_k}^{(m-1)}\right) &= \frac{p(\mathcal{D} \mid \tilde{\theta_k}^{(m-1)}, \mathcal{M})^{\beta_m} p(\tilde{\theta_k}^{(m-1)} \mid \mathcal{M})}{p(\mathcal{D} \mid \tilde{\theta_k}^{(m-1)}, \mathcal{M})^{\beta_{m-1}} p(\tilde{\theta_k}^{(m-1)} \mid \mathcal{M})} \\ &= p(\mathcal{D} \mid \tilde{\theta_k}^{(m-1)}, \mathcal{M})^{\beta_m - \beta_{m-1}} \end{aligned} \tag{3.6}$$

If N is chosen to be sufficiently large, then the samples $\tilde{\theta_k}^{(m-1)}, k = 1, \ldots, N$, are distributed according to the target PDF $f_{m-1}(\theta^{(m-1)})$ for the $(m-1)$th level that is given by normalizing $p(\mathcal{D} \mid \theta^{(m-1)}, \mathcal{M})^{\beta_{m-1}} p(\theta^{(m-1)} \mid \mathcal{M})$. Therefore, the expectation of $w(\theta^{(m-1)})$ under this PDF is approximated by the sample mean S_m of $w\left(\tilde{\theta_k}^{(m-1)}\right), k = 1, \ldots, N$:

$$\begin{aligned} S_m &= \frac{1}{N} \sum_{k=1}^{N} w\left(\tilde{\theta_k}^{(m-1)}\right) \\ &\approx \int \frac{p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_m} p(\theta \mid \mathcal{M})}{p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_{m-1}} p(\theta \mid \mathcal{M})} f_{m-1}(\theta) d\theta \\ &\approx \frac{\int p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_m} p(\theta \mid \mathcal{M}) d\theta}{\int p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_{m-1}} p(\theta \mid \mathcal{M}) d\theta} \end{aligned} \tag{3.7}$$

The products of these means of the re-sampling weights for all of the levels during one run of the TMCMC algorithm therefore gives the approximation

$$\begin{aligned} \prod_{m=1}^{M} S_m &\approx \frac{\int p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_M} p(\theta \mid \mathcal{M}) d\theta}{\int p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_0} p(\theta \mid \mathcal{M}) d\theta} \\ &= \frac{\int p(\mathcal{D} \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{M}) d\theta}{\int p(\theta \mid \mathcal{M}) d\theta} \\ &= \int p(\mathcal{D} \mid \theta, \mathcal{M}) p(\theta \mid \mathcal{M}) d\theta \\ &= p(\mathcal{D} \mid \mathcal{M}) \end{aligned} \tag{3.8}$$

Therefore, $\prod_{m=1}^{M} S_m$ is an estimator of the evidence, which is shown by Ching and Chen (2006) to be asymptotically unbiased.

3.4 Evaluating Robust Predictive PDFs Using Simulation

Equation 2.7 shows how a predictive PDF may be estimated for globally identifiable model classes of a system—but for systems with significant uncertainty in the parameters, this uncertainty must be taken into consideration using a more sophisticated approach. If the samples $\tilde{\theta}_k, k = 1 \dots K$, are simulated from the posterior distribution of θ, then the predictive PDF may be approximated by

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}) \approx \frac{1}{K} \sum_{k=1}^{K} p(X_{N+1}^M \mid \tilde{\theta}_k, \mathcal{M}) \tag{3.9}$$

3.4.1 Sensitivity of Predictive PDFs to Prior Distribution

The sensitivity of the posterior PDF and model-class selection results is demonstrated in Sections 2.2.4 and 2.3. Interestingly, stochastic simulation methods can be used to determine the sensitivity of the predictive PDFs to the choice of prior PDF (Beck, 2005).

Consider a set of model classes $\mathbf{M}$, consisting of N_m candidate models $\mathcal{M}_j, j = 1 \dots N_m$, plus a reference model class $\mathcal{M}_0$. Now assume that each model class differs only in the choice of the prior PDF over the vector of model parameters θ. The prior PDF for each model class is given by $p(\theta \mid \mathcal{M}_j) = \pi_j(\theta)$, with the reference model class having a prior distribution $p(\theta \mid \mathcal{M}_0) = \pi_0(\theta)$.

The models are defined over a set of predictive PDFs $p(X_1^M \mid \theta)$, where $X_1^M = \{x_1, \dots, x_M\}$. Let the data set $\mathcal{D}_N = \hat{X}_1^N$ consist of the first N measured vectors, then the posterior PDF on θ for the reference model class is given by

$$p(\theta \mid \mathcal{D}_N, \mathcal{M}_0) = \frac{p(\mathcal{D}_N \mid \theta)\pi_0(\theta)}{p(\mathcal{D}_N \mid \mathcal{M}_0)} \tag{3.10}$$

and the predictive PDF for X_{N+1}^M is given by

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}_0) = \int p(X_{N+1}^M \mid \mathcal{D}_N, \theta) p(\theta \mid \mathcal{D}_N, \mathcal{M}_0) d\theta \tag{3.11}$$

where the likelihood of the data is given by

$$p(X_{N+1}^M \mid \mathcal{D}_N, \theta) = \frac{p(\hat{X}_1^N, X_{N+1}^M \mid \theta)}{p(\hat{X}_1^N \mid \theta)} \tag{3.12}$$

Note that the predictive PDF for each candidate model class $\mathcal{M}_j$ can be expressed in terms of the reference model class as follows:

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}_j) = \frac{p(\mathcal{D}_N \mid \mathcal{M}_0)}{p(\mathcal{D}_N \mid \mathcal{M}_j)} \int p(X_{N+1}^M \mid \mathcal{D}_N, \theta) p(\theta \mid \mathcal{D}_N, \mathcal{M}_0) \frac{\pi_j(\theta)}{\pi_0(\theta)} d\theta \tag{3.13}$$

Now suppose that K samples are drawn from the posterior PDF of the reference model class $p(\theta \mid \mathcal{D}_N, \mathcal{M}_0)$, denoted $\hat{\theta}_k, k = 1 \ldots K$, then the approximation for Equation 3.13 can be made:

$$p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}_j) \approx \frac{p(\mathcal{D}_N \mid \mathcal{M}_0)}{p(\mathcal{D}_N \mid \mathcal{M}_j)} \frac{1}{N} \sum_{k=1}^{K} \frac{\pi_j(\hat{\theta}_k)}{\pi_0(\hat{\theta}_k)} p(X_{N+1}^M \mid \mathcal{D}_N, \hat{\theta}_k) \tag{3.14}$$

Similarly, the evidence for each model class in terms of the evidence of the reference model class can be approximated using the samples $\hat{\theta}_k$:

$$\begin{aligned} p(\mathcal{D}_N \mid \mathcal{M}_j) &= \int p(\mathcal{D}_N \mid \theta_j) \pi_j(\theta) d\theta \\ &= p(\mathcal{D}_N \mid \mathcal{M}_0) \int p(\theta \mid \mathcal{D}_N, \mathcal{M}_0) \frac{\pi_j(\theta)}{\pi_0(\theta)} d\theta \\ &\approx p(\mathcal{D}_N \mid \mathcal{M}_0) \frac{1}{N} \sum_{k=1}^{K} \frac{\pi_j(\hat{\theta}_k)}{\pi_0(\hat{\theta}_k)} \end{aligned} \tag{3.15}$$

Substituting this result into Equation 3.14,

$$\begin{aligned} p(X_{N+1}^M \mid \mathcal{D}_N, \mathcal{M}_j) &\approx \sum_{k=1}^{K} w_k p(X_{N+1}^M \mid \mathcal{D}_N, \hat{\theta}_k) \\ w_k &\equiv \frac{\pi_j(\hat{\theta}_k)}{\pi_0(\hat{\theta}_k)} \Bigg/ \sum_{i=1}^{K} \frac{\pi_j(\hat{\theta}_i)}{\pi_0(\hat{\theta}_i)} \end{aligned} \tag{3.16}$$

Since the likelihood function is independent of the choice of prior, variations in $\pi_j(\theta)$ only affect the weighting coefficients w_k, so the sensitivity of the predictive PDF

can be readily evaluated. It now also possible to compare the probabilities of the candidate models relative to the reference model:

$$\frac{P(\mathcal{M}_j \mid \mathcal{D}_N, \mathbf{M})}{P(\mathcal{M}_0 \mid \mathcal{D}_N, \mathbf{M})} = \frac{P(\mathcal{D}_N \mid \mathcal{M}_j)}{P(\mathcal{D}_N \mid \mathcal{M}_0)} \approx \frac{1}{N} \sum_{k=1}^{K} \frac{\pi_j(\hat{\theta}_k)}{\pi_0(\hat{\theta}_k)} \tag{3.17}$$

where the prior probabilities of the model classes, $P(\mathcal{M}_j \mid \mathbf{M})$, are chosen to be equal.

3.5 Conclusions

Several stochastic simulation methods are presented for use in Bayesian model updating. Each has specific advantages and disadvantages that will be highlighted in their application to various system-identification problems in the following chapters. Additionally, the samples generated using stochastic simulation can be used to perform Bayesian model-class selection and to approximate robust predictive PDFs, which allow predictions that are made with an updated model to directly incorporate modeling uncertainty.

Chapter 4

Ground Motion Attenuation Relations

A problem of much interest in earthquake engineering is the estimation of the ground motion that will be experienced at a given site due to an earthquake. Solving the full problem, which would involve the fault rupture process, propagation of the seismic waves through the Earth with reflection and refraction between strata, and the effects of the soil around the site, is far too complicated to perform on a routine basis, though studies involving sophisticated and very large finite-element models are being conducted (Aagaard, 1999; Krishnan et al., 2006). In practice, information recorded in previous earthquakes is used to make inferences about the expected ground motion at a site. This is done using regression equations, called ground motion attenuation equations, which use information such as the earthquake magnitude, distance, and site conditions to estimate an intensity measure for ground motion, such as the peak ground acceleration (PGA), or the spectral response at a specified period.

These attenuation equations are used in probabilistic loss estimation and, more recently, in efforts to develop seismic early warning methods (Cua, 2005; Grasso et al., 2005). The uncertainty associated with the regression parameters may contribute significantly to the overall uncertainty in the final decision variables generated by these methods (Grasso and Allen, 2005). Stochastic simulation methods allow for these uncertainties to be quantified. Additionally, Bayesian model-class selection can be used to choose the most appropriate form of the regression equation.

This problem is studied here to serve as a relatively simple example of the applications of Bayesian updating to system identification and model-class selection using real data.

4.1 Boore-Joyner-Fumal Attenuation Model

One well-known attenuation equation was first introduced by Joyner and Boore (1981), developed from strong-motion data recorded during the 1979 Imperial Valley, California, earthquake. Various refinements and modifications have been made since its introduction (Boore et al., 1993, 1997). This work will focus on the model presented in Boore et al. (1993). Sibilio et al. (2006) present an application of stochastic simulation methods to more recently developed ground-motion attenuation models. The model, which will be referred to as the Boore-Joyner-Fumal (BJF) equation for PGA, is given by

$$\begin{aligned} \log_{10}(PGA) &= b_1 + b_2(M-6) + b_3(M-6)^2 + b_4 R + b_5 \log_{10}(R) \\ &\quad + b_6 G_B + b_7 G_C + \epsilon \\ R &= \sqrt{d^2 + h^2} \end{aligned} \tag{4.1}$$

where M is the event magnitude; d is the distance to the surface projection of the causative fault; h is a fictitious depth parameter introduced to be representative of the depths of regional events; $G_B,\ G_C \in \{0, 1\}$ are binary soil classification parameters and ϵ is the prediction error. M, d, G_B, and G_C are input parameters for the attenuation model that define the seismic event and site soil conditions. The remaining parameters must be determined empirically from existing strong-motion data.

The regression analysis in Boore et al. (1993) was performed using Equation 4.1 and a database $\mathcal{D}$ of 271 strong-motion records from 20 earthquakes occurring in the western United States that ranged in magnitude from 5.2 to 7.7 and were recorded at distances ranging from 0 to 120 km.

The candidate model classes are defined by different combinations of the terms in

Model	Model Form for Mean $\log_{10}(PGA)$					
$\mathcal{M}_1$	$b_1 +$	$b_2(M-6)$	$+ b_3(M-6)^2$	$+ b_4 R$	$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_2$		$b_2(M-6)$	$+ b_3(M-6)^2$	$+ b_4 R$	$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_3$	$b_1 +$	$b_2(M-6)$		$+ b_4 R$	$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_4$	$b_1 +$	$b_2(M-6)$	$+ b_3(M-6)^2$		$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_5$		$b_2(M-6)$		$+ b_4 R$	$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_6$		$b_2(M-6)$	$+ b_3(M-6)^2$		$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_7$	$b_1 +$	$b_2(M-6)$			$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_8$		$b_2(M-6)$			$+ b_5 \log_{10}(R)$	$+ b_6 G_B + b_7 G_C$
$\mathcal{M}_9$	$b_1 +$	$b_2(M-6)$			$+ b_5 \log_{10}(R)$	

Table 4.1: Summary of model classes examined for the BJF attenuation relation.

Equation 4.1. Including every possible combination would result in a pool of $2^7 = 128$ candidate model classes. To reduce this number, the pool is restricted so that all model classes contain the linear magnitude term, the logarithmic R term, and the local soil-site-condition terms, except for one model class that excludes the site soil condition terms to study the effects of ignoring this information. The candidate model classes, $\mathcal{M}_j, j = 1, \ldots, 9$ are summarized in Table 4.1. Model class $\mathcal{M}_1$ includes all terms. Model classes $\mathcal{M}_2$ through $\mathcal{M}_4$ exclude one term, model classes $\mathcal{M}_5$ through $\mathcal{M}_7$ exclude two, and $\mathcal{M}_8$ excludes all three. Model class $\mathcal{M}_9$ includes only the constant term, the linear magnitude term, and the logarithmic R term.

The expressions in Table 4.1 establish the deterministic models for the system. To create probability models, as discussed in Section 2.1, the model prediction error ϵ is assumed to be zero-mean Gaussian white noise, implying stochastic independence between predictions for different event magnitudes, distances, and site soil conditions. The resulting likelihood function is of the form

$$p(\mathcal{D} \mid \mathcal{M}, \theta) = \sum_{i=1}^{N_e} \sum_{j=1}^{N_{r,i}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{1}{2\sigma^2} (\log_{10}[PGA(i,j)] - Y(i,j \mid \theta))^2 \right) \qquad (4.2)$$

where N_e is the number of seismic events comprising the data $\mathcal{D}$, $N_{r,i}$ is the number of records in the data for the ith event, $PGA(i,j)$ is the recorded peak ground acceleration for the jth record of the ith event, and $Y(i,j \mid \theta)$ is the PGA predicted by Equation 4.1 for the same record, given the parameter vector θ, containing the

regression coefficients $b = [b_1, b_2, b_3, b_4, b_5, b_6, b_7]$ and the fictitious depth parameter h.

The final component needed to completely define each Bayesian model class, $\mathcal{M}_j, j = 1, \ldots, 9$, is the prior PDF over the model parameters. The prior PDF for each component of b included in a given model class is taken to be an independent Gaussian distribution with zero mean and a standard deviation of $\sigma_b = 10$, which is considered to be a fairly broad prior, as shown in Figure 4.1(a). The prior PDF for h for all model classes is a lognormal distribution with logarithmic mean ln[10.43], which is the log of the mean of the available estimates for the depths of the seismic events comprising data set $\mathcal{D}$, and a logarithmic standard deviation of 0.5, resulting in the PDF shown in Figure 4.1(b). For the prediction-error variance σ^2, an inverse gamma distribution is used for all model classes:

$$p(\sigma^2 \mid \mathcal{M}_j) = \frac{\beta^\alpha}{\Gamma(\alpha)(\sigma^2)^{\alpha+1}} \exp\left[-\frac{\beta}{\sigma^2}\right] \tag{4.3}$$

where the shape parameter α and the scaling parameter β are defined by the user and $\Gamma(\alpha)$ is the gamma function evaluated at α. For convenience, the inverse gamma distribution will be denoted $IG(\alpha, \beta)$. The inverse gamma distribution is a *conditionally conjugate* distribution, which is a property that will be shown in the next section to be very useful in improving the efficiency of the Gibbs sampler. The parameters chosen for the prior PDF are $\alpha = 3$, chosen to produce what is considered a reasonable shape for the prior PDF, and $\beta = 0.1165$, which is the variance of the measured values for $\log_{10}(PGA)$ in the data $\mathcal{D}$ and thus is believed to scale the distribution appropriately for this problem. The resulting PDF is shown in Figure 4.1(c).

With the Bayesian model classes now defined, Bayesian updating is performed to obtain the posterior PDF for the regression coefficients $b = [b_1, b_2, b_3, b_4, b_5, b_6, b_7]$, the fictitious depth parameter h, and the prediction-error variance σ^2.

4.1.1 Estimation and Selection with Gibbs Sampler

The application of Gibbs sampling is as follows. In the first stage, samples for the fictitious depth h and prediction-error variance σ^2 are drawn from the prior PDFs,

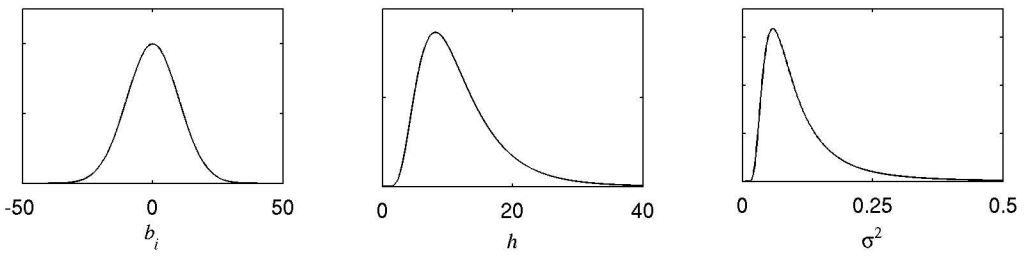


Figure 4.1: Prior PDFs for the BJF model parameters, which include (a) the regression coefficients b_i, (b) the fictitious depth parameter h, and (c) the prediction-error variance σ^2.

denoted $\tilde{h}^{(0)}$ and $(\tilde{\sigma}^{(0)})^2$, respectively. The goal for this stage is to draw a sample for b from the conditional PDF $p(b \mid \tilde{h}^{(0)}, (\tilde{\sigma}^{(0)})^2, \mathcal{D}, \mathcal{M})$. The resulting likelihood function for the conditional PDF is linear in b, and it can be shown fast sampling from this PDF is possible. Consider a linear system

$$\hat{Y} = AX + E \tag{4.4}$$

where $\hat{Y}$ is a vector of observations and A is a fixed matrix. X is a vector of uncertain variables, which has a Normal prior PDF, $N(X_0, \Sigma_0)$. The prediction error E is zero-mean and Gaussianly distributed with a covariance matrix Σ, that is $E \sim N(0, \Sigma)$. Gelfand et al. (1990) showed that the mean and covariance matrix of the Gaussian distribution of X conditioned on $\hat{Y}$ are given by

$$\begin{aligned} E(X \mid \hat{Y}) &= X_0 + \Sigma_0 A^T (A\Sigma_0 A^T + \Sigma)^{-1} (\hat{Y} - AX_0) \\ Cov(X \mid \hat{Y}) &= \Sigma_0 - \Sigma_0 A^T A^T (A\Sigma_0 A^T + \Sigma)^{-1} A\Sigma_0 \end{aligned} \tag{4.5}$$

Sampling of $p(X \mid \hat{Y})$ is then trivial to perform. Appendix B shows in detail how these expressions are derived. Rather than considering the entire data set at one time, it is computationally more efficient to sequentially update with the data recorded during each of the N_e earthquakes comprising the data set, as outlined in Section 2.1.3 because it avoids inversion of large matrices in Equation 4.5. For updating with the $N_{r,1}$ records from the first earthquake in the data set, the vector $\hat{Y} = PGA_1$,

the vector of measured PGAs for the first earthquake, while X equals the vector of the components of b included the given model class, and X_0 is the vector of mean values for the prior PDF over those components of b, chosen to be a vector of zeros in Section 4.1. The matrices A, Σ, and Σ_0 depend on the selected model class. For example, for model class $\mathcal{M}_7$ the matrix A is given by

$$\begin{aligned} A &= \left[\begin{array}{cccc} \mathbf{1}_{N_{r,1}\times 1} & (M_1 - 6)\mathbf{1}_{N_{r,1}\times 1} & R_1^* & G_{B,1} \quad G_{C,1} \end{array} \right] \\ R_1^* &= [\log_{10}(R_{1,j}),\ j = 1, \ldots, N_{r,1}]^T \\ R_{1,j} &= \sqrt{d_{1,j}^2 + (\tilde{h}^{(0)})^2} \end{aligned} \tag{4.6}$$

where $\mathbf{1}_{N_{r,1}\times 1}$ is a $N_{r,1}\times 1$ vector of ones, $d_{1,j}$ is the distance from the surface projection of the causitive fault to the jth measurement location for the first earthquake in the data set, and $G_{B,1}$ and $G_{C,1}$ are the vectors of site soil classification parameters for the measurement sites. The matrices Σ and Σ_0 for model class $\mathcal{M}_7$ are given by:

$$\Sigma = (\tilde{\sigma}^{(0)})^2 I_{N_{r,i}\times N_{r,i}} \tag{4.7}$$

$$\Sigma_0 = \sigma_b^2 I_{5\times 5} \tag{4.8}$$

where σ_b is the standard deviation of the prior PDF for each component of b, chosen to be 10 in Section 4.1. Using Equation 4.5, the mean and covariance matrix of this conditional PDF are obtained, and denoted μ_1 and Σ_1, respectively. Updating with the data from the second earthquake in the data set is performed in a similar manner; however, in this case, the vector $X_0 = \mu_1$ and the matrix $\Sigma_0 = \Sigma_1$. The process is repeated until the final mean and covariance matrix, μ_{N_e} and Σ_{N_e}, are obtained and used to generate the sample $\tilde{b}^{(1)}$, from a Gaussian PDF, $N(\mu_{N_e}, \Sigma_{N_e})$.

In the second stage, the values for b and h are fixed (at $\tilde{b}^{(1)}$ and $\tilde{h}^{(0)}$, respectively), and the conditionally distributed sample for σ^2 is drawn from $p(\sigma^2 \mid \tilde{b}^{(1)}, \tilde{h}^{(0)}, \mathcal{D}, \mathcal{M})$. Recall that the chosen prior PDF is an inverse gamma distribution, $IG(\alpha, \beta)$, which is a conditionally conjugate prior. That is, the conditional PDF is of the same form as the prior PDF. It can be shown that the conditional PDF is proportional to an

inverse gamma distribution:

$$p(\sigma^2 \mid \tilde{b}^{(1)}, \tilde{h}^{(0)}, \mathcal{D}, \mathcal{M}) \propto IG\left(\alpha + \frac{1}{2}\sum_{i=1}^{N_e} N_{r,i},\ \beta + \sum_{i=1}^{N_e}\sum_{j=1}^{N_{r,i}} \delta_{i,j}^2\right) \tag{4.9}$$
$$\delta_{i,j} = PGA(i,j) - Y(i,j \mid \tilde{b}^{(1)}, \tilde{h}^{(0)}$$

Sampling from this distribution to obtain $(\tilde{\sigma}^{(1)})^2$ is therefore a simple procedure. Appendix B gives the proof for this result, which is only true when the errors are uncorrelated, (i.e. the covariance matrix for the prediction errors is constrained to be diagonal).

The final step, drawing a sample for h from the conditional posterior PDF $p(h \mid \tilde{b}^{(1)}, (\tilde{\sigma}^{(1)})^2, \mathcal{D}, \mathcal{M})$ is more complicated, since h enters into the regression equation as a non-linear term, preventing the direct calculation of the mean and covariance matrix in Equation 4.5. Two approaches were used for drawing from this conditional distribution.

For the first approach, the M-H algorithm was used to generate samples from the conditional PDF. Since only one parameter is involved, the computational cost is reasonable. After discarding a burn-in period of 10 samples, $h^{(1)}$ is drawn from the remaining samples with each sample equally likely to be selected. It was found that between 200–300 samples from the M-H algorithm are necessary for convergence to the conditional PDF for h.

The second approach uses an analytical approximation of the conditional posteior PDF for h. The logarithm of this conditional PDF is proportional to the function $f(h)$, defined as

$$f(h) = p(\mathcal{D} \mid h, \tilde{b}^{(1)}, (\tilde{\sigma}^{(1)})^2, \mathcal{M})\ p(h \mid \mathcal{M}) \tag{4.10}$$

$f(h)$ is numerically optimized to determine the most probable value $\hat{h}$. An analytical expression for the curvature of $f(h)$ is evaluated at $\hat{h}$ and used to form a lognormal approximation to the posterior PDF, similar in concept to the Gaussian approximation in Equation 2.20. This lognormal PDF is used to generate the sample $h^{(1)}$. Sampling using this method is much faster than the embedded M-H method. While

the conditional PDF should only be well approximated by the lognormal distribution in the neighborhood of $\hat{h}$, in practice the results are very similar to those generated using embedded M-H to sample the conditional PDF.

Samples generated using the Gibbs sampler (GS) approach for model class $\mathcal{M}_1$ are shown in Figure 4.2. A burn-in period of 30 samples was selected based on observation of the initial runs of the sampler. The evolution of the mean and standard deviation as more samples are added is shown in Figure 4.3. Based on these results, a chain length of 600 samples is selected.

Three parallel chains of 600 samples (with a 30 sample burn-in period) were simulated for each model class. The estimated most probable parameter values and standard deviations of the marginal posterior PDFs for each parameter are shown in Table 4.2 for samples generated with the GS algorithm using an embedded Metropolis-Hasting sampler to sample the conditional PDF for h. The same quantities, estimated for samples generated with the GS algorithm using the asymptotic approximation for the conditional PDF of h are shown in Table 4.3.

The evidence for each model class was estimated using the post burn-in samples outlined in Equations 3.3 through 3.5 in Section 3.3 and used to determine the probability of each model class. The results of model-class selection are shown in Tables 4.2 and 4.3. Clearly, model class $\mathcal{M}_8$ is the most probable one based on the data $\mathcal{D}$, by a wide margin.

Figure 4.4 shows normalized histograms of the samples generated for the regression coefficients and fictitious depth parameter for model class $\mathcal{M}_1$ using GS with an embedded M-H sampler. The prior PDFs for each parameter are plotted, along with the optimal values (obtained by optimizing the posterior PDF). Note that there is good agreement between the optimal values and the high-probability regions indicated by the samples. Also, note that for the regression coefficients, the prior PDF is barely visible, indicating that the posterior is far more peaked than the prior PDF. This is true even though Table 4.2 indicates that $\mathcal{M}_1$ has larger uncertainties for the parameter estimates (broader posterior PDFs for the parameters) than the other model classes, which is expected since the same data is being used to identify a greater

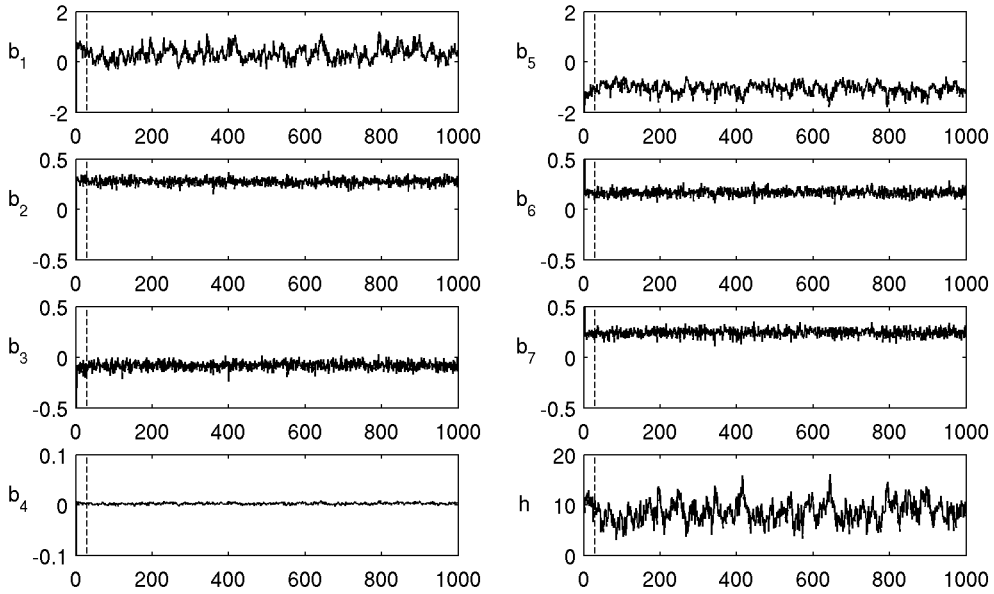


Figure 4.2: Samples of BJF regression coefficients for model class $\mathcal{M}_1$ generated with the GS approach. The index n denotes the step number of the Markov chain. The dashed line represents the end of the burn-in period at $n = 30$.

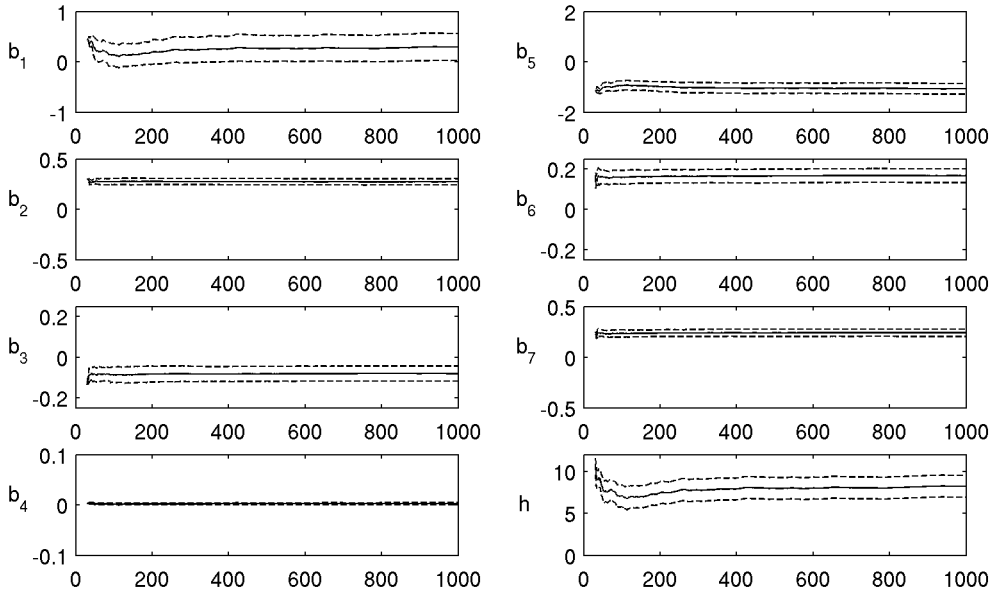


Figure 4.3: Mean and standard deviation of samples in Figure 4.2 plotted against the number of samples included (excluding a 30 sample burn-in period). The solid line is the sample mean, and the dashed lines represent the mean plus and minus one standard deviation.

Mdl.	b_1	b_2	b_3	b_4	b_5	b_6	b_7	h	σ	Prob
$\mathcal{M}_1$	0.192	0.275	-0.081	0.002	-0.992	0.169	0.241	7.90	0.193	0.0
	(0.269)	(0.032)	(0.037)	(0.002)	(0.210)	(0.036)	(0.038)	(2.16)	(0.009)	
$\mathcal{M}_2$		0.273	-0.080	0.001	-0.839	0.174	0.248	6.32	0.196	0.0
		(0.032)	(0.038)	(0.001)	(0.038)	(0.035)	(0.037)	(1.00)	(0.009)	
$\mathcal{M}_3$	0.195	0.226		0.003	-1.025	0.174	0.263	7.92	0.197	0.0
	(0.270)	(0.022)		(0.003)	(0.212)	(0.035)	(0.036)	(2.08)	(0.009)	
$\mathcal{M}_4$	-0.067	0.275	-0.075		-0.770	0.164	0.240	6.08	0.196	0.1
	(0.100)	(0.033)	(0.038)		(0.062)	(0.035)	(0.035)	(1.51)	(0.009)	
$\mathcal{M}_5$		0.223		0.001	-0.865	0.179	0.269	6.44	0.197	0.4
		(0.022)		(0.001)	(0.035)	(0.033)	(0.034)	(0.96)	(0.009)	
$\mathcal{M}_6$		0.276	-0.072		-0.801	0.155	0.232	6.87	0.197	4.0
		(0.032)	(0.037)		(0.026)	(0.032)	(0.034)	(0.98)	(0.009)	
$\mathcal{M}_7$	-0.052	0.226			-0.809	0.172	0.263	6.38	0.197	2.0
	(0.095)	(0.022)			(0.056)	(0.035)	(0.036)	(1.31)	(0.009)	
$\mathcal{M}_8$		0.230			-0.834	0.164	0.256	6.78	0.197	93.4
		(0.022)			(0.020)	(0.031)	(0.032)	(0.97)	(0.008)	
$\mathcal{M}_9$	0.253	0.226			-0.890			6.92	0.216	0.0
	(0.114)	(0.025)			(0.072)			(1.49)	(0.009)	

Table 4.2: Most probable values of the marginal posterior PDFs for each parameter (with standard deviations in parentheses) and model-class selection results for ground motion attenuation based on samples using the GS algorithm with an embedded M-H sampler for the conditional PDF of h.

Mdl.	b_1	b_2	b_3	b_4	b_5	b_6	b_7	h	σ	Prob
$\mathcal{M}_1$	0.235	0.274	-0.082	0.003	-1.029	0.167	0.240	7.94	0.196	0.0
	(0.280)	(0.032)	(0.037)	(0.002)	(0.218)	(0.035)	(0.036)	(2.19)	(0.009)	
$\mathcal{M}_2$		0.272	-0.079	0.001	-0.835	0.172	0.248	6.41	0.197	0.0
		(0.032)	(0.037)	(0.001)	(0.038)	(0.034)	(0.036)	(1.04)	(0.009)	
$\mathcal{M}_3$	0.289	0.223		0.003	-1.028	0.175	0.264	7.90	0.198	0.0
	(0.275)	(0.023)		(0.002)	(0.215)	(0.035)	(0.036)	(2.18)	(0.009)	
$\mathcal{M}_4$	-0.036	0.274	-0.074		-0.782	0.163	0.240	6.23	0.197	0.1
	(0.128)	(0.033)	(0.038)		(0.073)	(0.034)	(0.037)	(1.62)	(0.009)	
$\mathcal{M}_5$		0.223		0.001	-0.863	0.178	0.269	6.51	0.198	0.4
		(0.023)		(0.001)	(0.036)	(0.034)	(0.035)	(1.02)	(0.009)	
$\mathcal{M}_6$		0.276	-0.073		0.802	0.156	0.232	6.83	0.197	3.6
		(0.032)	(0.038)		(0.026)	(0.032)	(0.035)	(1.01)	(0.009)	
$\mathcal{M}_7$	-0.028	0.227			-0.819	0.170	0.260	6.40	0.198	1.7
	(0.102)	(0.023)			(0.059)	(0.034)	(0.034)	(1.45)	(0.009)	
$\mathcal{M}_8$		0.229			-0.834	0.165	0.255	6.88	0.198	94.2
		(0.021)			(0.020)	(0.032)	(0.033)	(1.02)	(0.009)	
$\mathcal{M}_9$	0.289	0.228			-0.907			7.12	0.216	0.0
	(0.107)	(0.025)			(0.067)			(1.59)	(0.009)	

Table 4.3: Most probable values of the marginal posterior PDFs for each parameter (with standard deviations in parentheses) and model-class selection results for ground motion attenuation based on samples using the GS algorithm with an analytical approximation for the conditional PDF of h.

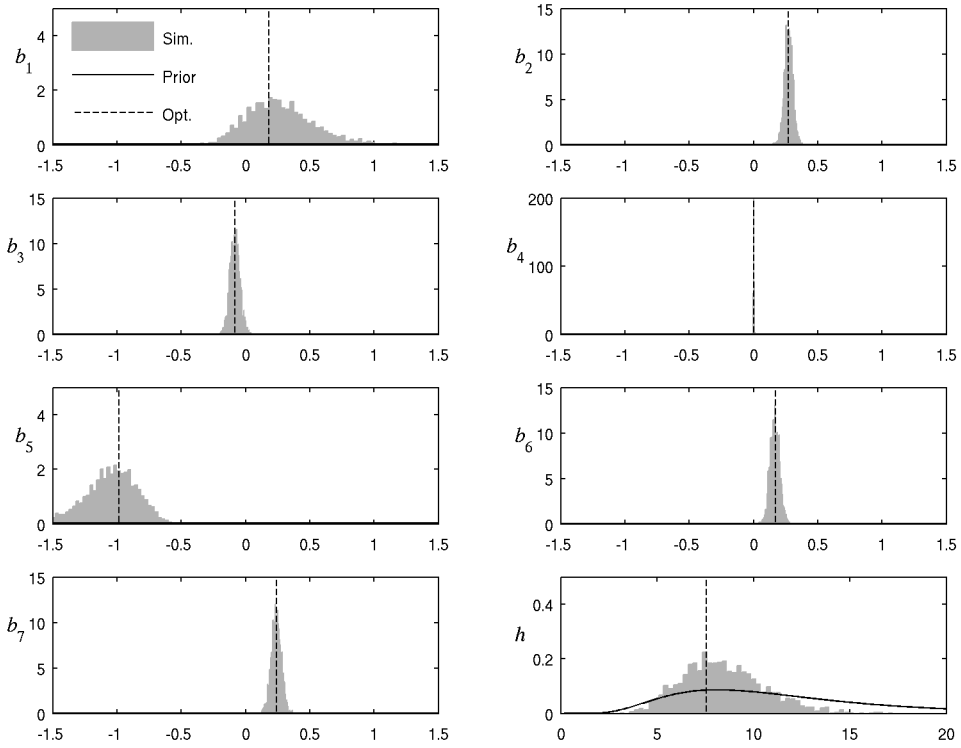


Figure 4.4: Normalized histograms for samples generated for the parameters of model class $\mathcal{M}_1$ using the GS algorithm with an embedded M-H sampler for the conditional PDF for h, plotted with the prior PDFs for each parameter and the optimal estimate for the parameter.

number of parameters. However, the marginal posterior PDF for the fictitious depth parameter h is still fairly broad after updating for all model classes.

4.1.2 Estimation and Selection with TMCMC

The TMCMC algorithm is also used to perform updating for the candidate model classes. Five chains of 1500 samples were generated for each model class. Convergence to the target PDF typically required between 20–30 levels, depending on the complexity of the model. Table 4.4 gives the estimated most probable model parameter values and standard deviations of the marginal PDFs for each parameter obtained from the samples.

Mdl.	b_1	b_2	b_3	b_4	b_5	b_6	b_7	h	σ	Prob.
$\mathcal{M}_1$	0.283	0.276	-0.083	0.003	-1.070	0.172	0.245	8.44	0.202	0.0
	(0.296)	(0.035)	(0.042)	(0.002)	(0.229)	(0.037)	(0.039)	(2.27)	(0.009)	
$\mathcal{M}_2$		0.269	-0.078	0.001	-0.838	0.174	0.248	6.26	0.200	0.0
		(0.034)	(0.039)	(0.001)	(0.040)	(0.036)	(0.038)	(1.02)	(0.009)	
$\mathcal{M}_3$	0.307	0.226		0.003	-1.115	0.173	0.266	9.05	0.204	0.0
	(0.333)	(0.025)		(0.002)	(0.255)	(0.037)	(0.039)	(2.39)	(0.009)	
$\mathcal{M}_4$	-0.075	0.272	-0.074		-0.763	0.161	0.240	5.95	0.202	0.0
	(0.100)	(0.034)	(0.040)		(0.062)	(0.037)	(0.037)	(1.03)	(0.009)	
$\mathcal{M}_5$		0.221		0.001	-0.864	0.178	0.270	6.37	0.203	0.0
		(0.024)		(0.001)	(0.038)	(0.037)	(0.037)	(1.03)	(0.009)	
$\mathcal{M}_6$		0.275	-0.074		-0.800	0.155	0.229	6.77	0.200	1.4
		(0.033)	(0.038)		(0.027)	(0.033)	(0.036)	(1.04)	(0.009)	
$\mathcal{M}_7$	-0.051	0.228			-0.809	0.171	0.262	6.23	0.202	0.3
	(0.107)	(0.023)			(0.062)	(0.037)	(0.038)	(1.46)	(0.009)	
$\mathcal{M}_8$		0.230			-0.834	0.161	0.253	6.71	0.201	98.3
		(0.023)			(0.022)	(0.034)	(0.034)	(1.01)	(0.009)	
$\mathcal{M}_9$	0.252	0.227			-0.889			6.78	0.221	0.0
	(0.111)	(0.025)			(0.070)			(1.61)	(0.010)	

Table 4.4: Most probable values of the marginal posterior PDFs for each parameter (with standard deviations in parentheses) and model-class selection results for ground motion attenuation based on samples using Transitional Markov Chain Monte Carlo.

For each run of the TMCMC sampler, the evidence for each model class is estimated as outlined in Equations 3.6 through 3.8. The mean values of the five estimates of the evidence for each model class were used to calculate the model class probabilities, which are also shown in Table 4.4.

4.1.3 Comparison of Simulation Methods

For purposes of comparison, the most probable values based on the posterior distribution were estimated from direct numerical optimization of the log of the posterior PDF. The Hessian matrix of the log of the posterior is analytically derived from Equation 4.2 and then evaluated using the most probable values of the parameters. The standard deviations of the model parameters were then estimated from the covariance matrix for the Gaussian approximation of the posterior PDF, which is given by the negative inverse of the Hessian matrix, as in Equation 2.20. The results are listed in Table 4.5. Estimates of the most probable values and standard deviations of each parameter, as shown in Tables 4.2 through 4.4 for the different methods are fairly

Mdl.	b_1	b_2	b_3	b_4	b_5	b_6	b_7	h	σ	Prob.
$\mathcal{M}_1$	0.182	0.273	-0.081	0.002	-0.986	0.170	0.242	7.57	0.193	0.0
	(0.272)	(0.032)	(0.037)	(0.002)	(0.216)	(0.034)	(0.036)	(2.28)	(0.008)	
$\mathcal{M}_2$		0.271	-0.079	-0.001	-0.841	0.173	0.247	6.31	0.193	0.0
		(0.032)	(0.037)	(0.001)	(0.037)	(0.034)	(0.036)	(0.98)	(0.008)	
$\mathcal{M}_3$	0.157	0.223		0.002	-0.993	0.176	0.264	7.50	0.195	0.0
	(0.274)	(0.022)		(0.002)	(0.218)	(0.034)	(0.035)	(2.27)	(0.009)	
$\mathcal{M}_4$	-0.075	0.273	-0.075		-0.760	0.165	0.241	6.02	0.194	0.0
	(0.100)	(0.032)	(0.037)		(0.062)	(0.034)	(0.036)	(1.50)	(0.008)	
$\mathcal{M}_5$		0.222		0.001	-0.868	0.178	0.268	6.43	0.195	0.0
		(0.022)		(0.001)	(0.035)	(0.034)	(0.034)	(0.97)	(0.009)	
$\mathcal{M}_6$		0.277	-0.072		-0.802	0.155	0.230	6.81	0.194	2.8
		(0.032)	(0.037)		(0.025)	(0.031)	(0.034)	(1.02)	(0.008)	
$\mathcal{M}_7$	-0.063	0.226			-0.800	0.171	0.262	6.19	0.195	1.2
	(0.103)	(0.022)			(0.060)	(0.034)	(0.035)	(1.49)	(0.009)	
$\mathcal{M}_8$		0.230			-0.834	0.163	0.252	6.83	0.195	96.0
		(0.021)			(0.020)	(0.031)	(0.032)	(0.99)	(0.009)	
$\mathcal{M}_9$	0.243	0.227			-0.881			6.77	0.214	0.0
	(0.108)	(0.024)			(0.068)			(1.63)	(0.009)	

Table 4.5: Most probable parameter values (with standard deviations in parentheses) and model-class selection results for ground motion attenuation using asymptotic approach.

consistent, especially since some errors are expected due to the differences that arise from maximizing the joint posterior PDF for the "optimal" results and maximizing the individual marginal posterior PDFs for the simulations results (see Section 3.2). The agreement between the direct optimization and simulation methods can clearly be seen in Figure 4.4, which shows the optimal parameter values for $\mathcal{M}_1$ plotted against histograms (estimated marginal PDFs) for each parameter and in Figure 4.5, which shows the optimal parameter values for $\mathcal{M}_7$ plotted with the samples generated for $\mathcal{M}_7$ using the GS approach (with a Metropolis-Hastings update for h) projected on a series of 2-D sub-spaces of the parameter space.

The results from model-class selection are similarly consistent, with the ordering of the three most probable model classes the same for all methods. Though there are some differences in the probabilities estimated for each model class, all methods clearly indicate that the preferred model is $\mathcal{M}_8$ with a probability of over 90% based on the data $\mathcal{D}$. The omission of the quadratic term in magnitude and the linear term in R is unsurprising, given that the associated regression coefficients are very nearly zero when those terms are included, with very small uncertainties, indicating

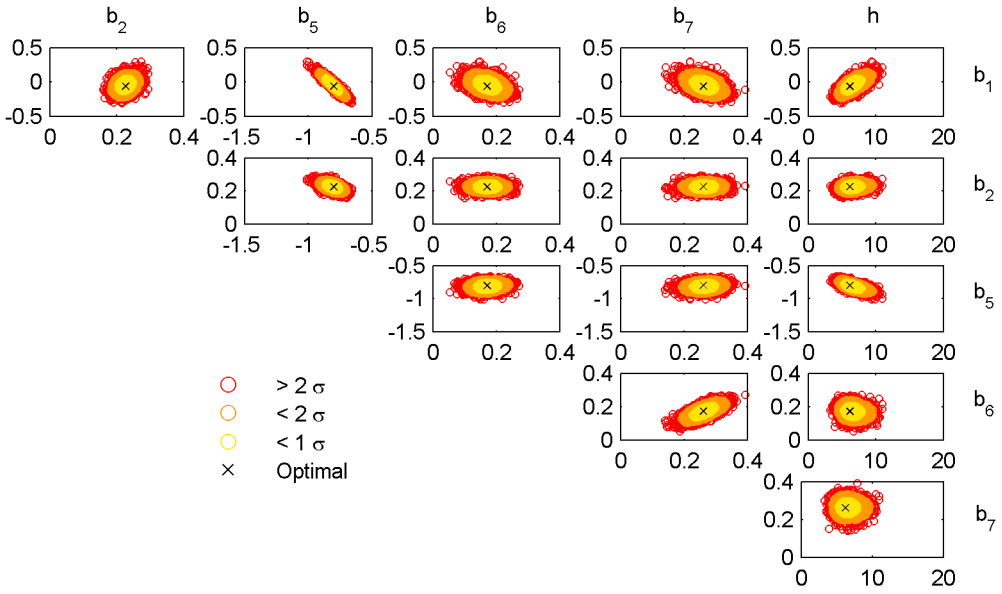


Figure 4.5: Samples of the BJF regression coefficents included in $\mathcal{M}_7$ projected onto a series of 2-D sub-spaces of the parameter space. Color indicates the distance of the sample from the mean. The results of the direct numerical optimization of the posterior are plotted for comparison

that these terms have little effect on the data fit and extract a large amount of information from the data. The constant term b_1 is, in some models, large enough to have a significant effect on the data. However, the uncertainty is quite large, and the loss in accuracy can be compensated to some degree by other parameters. Figure 4.5 shows significant interaction of b_1 with both the coefficient b_5, which is associated with the logarithmic R term, and the depth parameter h, and that fixing b_1 at zero still leaves the parameters in a high-probability region.

The model-class selection results also indicate the importance of the site soil conditions. Though model class $\mathcal{M}_9$ has a smaller number of uncertain parameters than the other model classes, the simplicity is outweighed by poorer data fitting ability.

4.1.4 Comparison of Model Class Predictions

It is also of interest to compare the differences between the predictions given by different model classes. Figure 4.7 shows the estimated posterior mean of the predicted

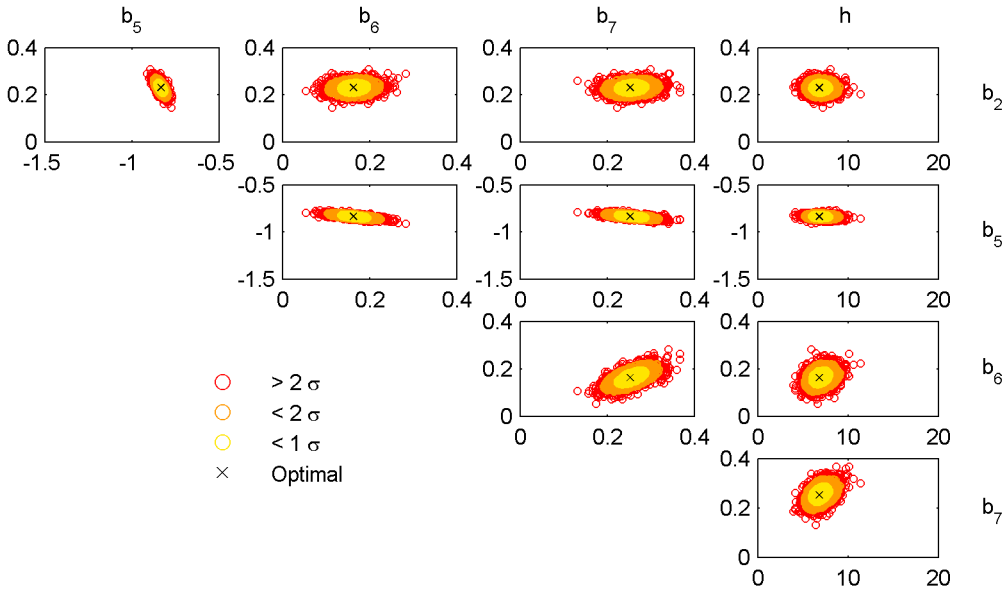


Figure 4.6: Samples of the BJF regression coefficents included in $\mathcal{M}_8$ projected onto a series of 2-D sub-spaces of the parameter space. Color indicates the distance of the sample from the mean. The results of the direct numerical optimization of the posterior are plotted for comparison

relationship between site distance and PGA for the three most probable model classes (in order of probability, $\mathcal{M}_8$, $\mathcal{M}_6$, and $\mathcal{M}_7$) for soil sites (site classes B and C) for events of magnitude 5.5, 6.5 and 7.5. Note that while model class $\mathcal{M}_8$, the most probable model class, and model class $\mathcal{M}_7$, which has the same general form as the model identified by Boore et al. (1993), give nearly identifical curves, model class $\mathcal{M}_6$, which includes a quadratic magnitude term that is not included in the other two model classes, gives a quite different response. Though model class $\mathcal{M}_7$ appears to be relatively close to the most probable model class, the model-class selection results shown in Tables 4.2 through 4.5 show that model class $\mathcal{M}_6$ is at least twice as probable as model class $\mathcal{M}_7$. The measured PGA data for soil sites is also shown in Figure 4.7, showing there is considerable scatter in the data.

Figure 4.8 shows a similar comparison of the prediction for each model class for the relationship between magnitude and PGA for soil sites at distances of 5 km, 20 km and 50 km. Again model class $\mathcal{M}_8$ and $\mathcal{M}_7$ give very similar predictions, while model class $\mathcal{M}_6$ differs significantly. The effect of the quadratic magnitude term (intended

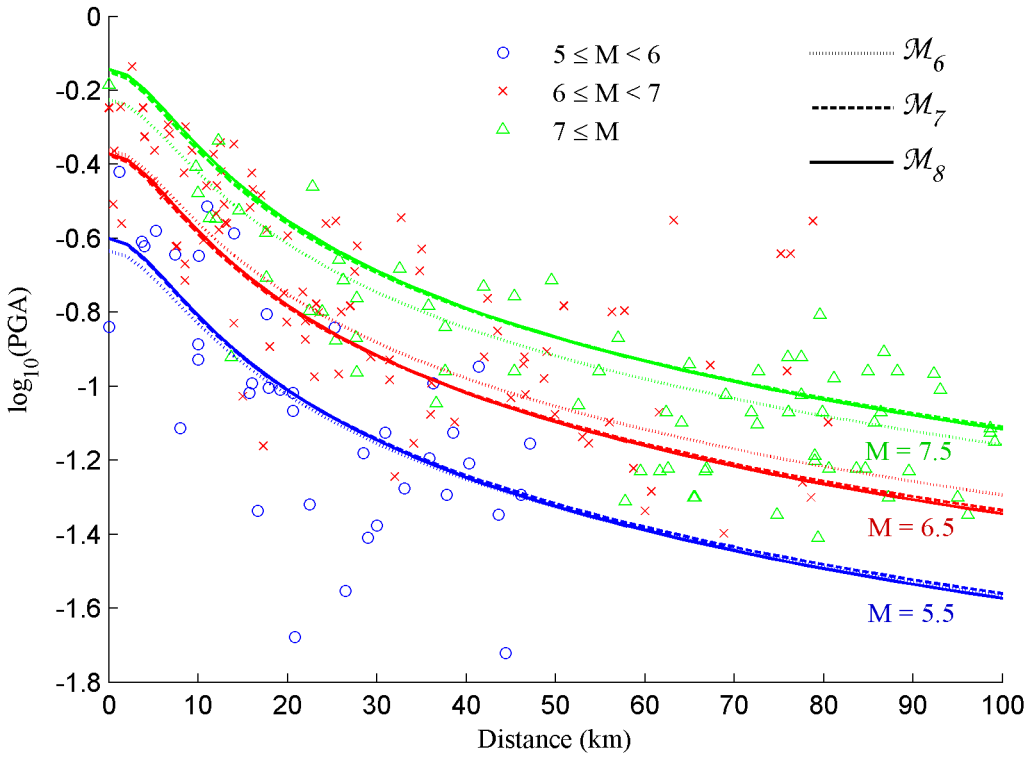


Figure 4.7: Predicted distance-PGA curves for the three most probable model classes at magnitudes of 5.5, 6.5 and 7.5 for soil sites (site classes B and C), plotted with PGA data for soil sites.

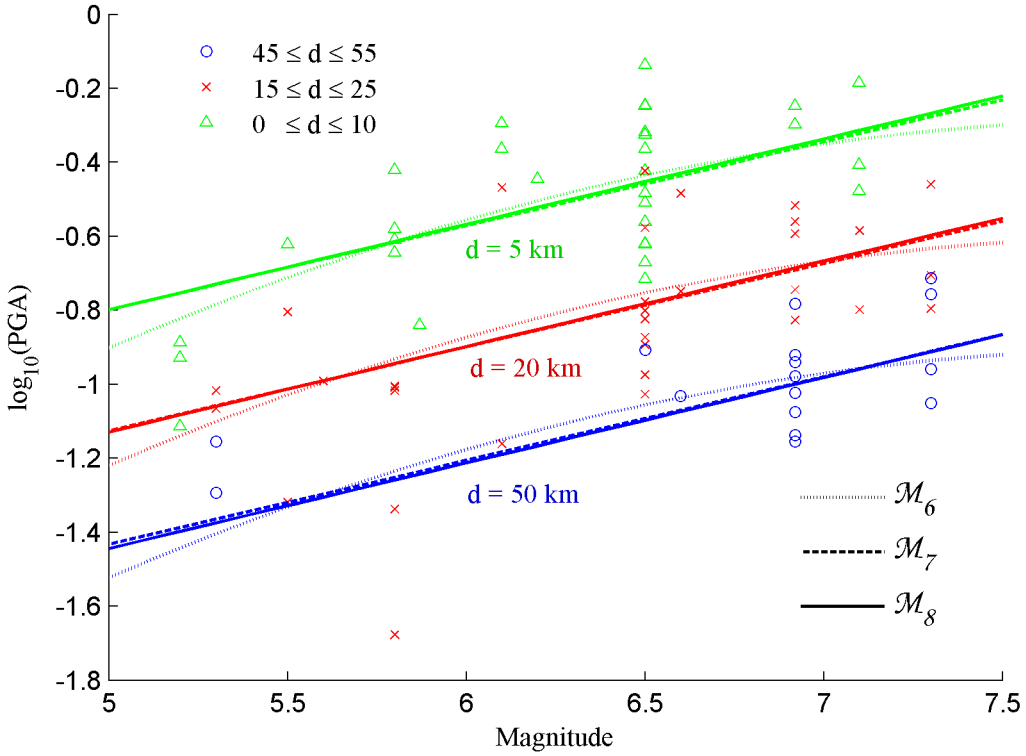


Figure 4.8: Predicted magnitude-PGA curves for the three most probable model classes at distances of 5, 20 and 50 km for soil sites (site classes B and C), plotted with PGA data for soil sites.

to represent magnitude saturation) in $\mathcal{M}_6$ is immediately evident.

Figure 4.8 also illustrates an important issue in the use of predictive models; specifically the reliability of predictions in cases where little or no data is available. The similar values for prediction-error variance, shown in Tables 4.2 through 4.5, indicate that there is little difference in the data-fitting ability of model classes $\mathcal{M}_6$, $\mathcal{M}_7$, and $\mathcal{M}_8$. However, it is clear from Figure 4.8 that the model class predictions for earthquakes of magnitude greater than 7.5 will be substantially different for model class $\mathcal{M}_6$. If data for larger magnitude events were used for model updating and model-class selection, it would very probably alter the model-class probabilities significantly, particularly given recent work which provides strong evidence that the relationship between PGA and distance is magnitude-dependent (Cua, 2005).

4.2 Conclusions

Bayesian updating was performed on the Boore-Joyner-Fumal attenuation relation using two different stochastic simulation methods. It was shown that the Gibbs sampler (GS) algorithm can perform efficient sampling of the conditional distribution for linear models, which can result in computational savings even if the model ouput is not linear in all components. This property makes the GS algorithm potentially very useful in dealing with models with large numbers of parameters, as will be demonstrated in the next chapter. The TMCMC was succesfully implemented to give samples of the posterior PDF that are reasonably consistent with the results from the GS algorithm and the most probable model estimated using numerical optimization of the posterior PDF.

Model-class selection was performed, demonstrating how Bayesian model-class selection balances fitting of the available data with the principle of parsimonious modeling, as the results favored model classes that eliminated unimportant or redundant terms. Estimates of the evidence for each model class obtained using samples generated using the GS algorithm are consistent with estimates based on the asymptotic approximation of the evidence, a technique that has been demonstrated to be effective for dynamical systems (Beck and Yuen, 2004). Estimates of the evidence obtained using samples generated with TMCMC were qualitatively consistent with the other methods, but discrepancies in the relative weighting of the models suggest the need for further investigation into the evidence estimation aspect of the TMCMC algorithm.

Chapter 5

Structural Health Monitoring

In civil engineering, structural health monitoring (SHM) typically refers to techniques for the non-destructive evaluation of the condition of civil structures (Natke and Yao, 1988; Doebling et al., 1996; Bernal et al., 2002). One particular approach involves comparing the posterior PDFs of linear structural model parameters updated using modal properties obtained from ambient vibration data before and after potentially damaging events to compute probabilities of sub-structure damage (Vanik et al., 2000; Yuen et al., 2004; Ching and Beck, 2004).

It will be shown in this chapter that stochastic simulation methods may be efficiently applied to this problem to obtain probability distributions for damage based on data from the undamaged and potentially damaged system, which can be combined to obtain a probabilistic description of the location and severity of damage. The simulation techniques used allow for updating of structural parameters associated with degrees of freedom that are not directly observed in the data and can handle problems with large numbers of uncertain parameters. The work presented in this chapter was performed in collaboration with Dr. Jianye Ching and elements of it are presented in Ching et al. (2005, 2006).

5.1 Linear Model Updating with Incomplete Modal Data

The dynamic response of a structure under strong earthquake shaking, particularly in the case where damage occurs, is highly non-linear. However, when modal parameters are extracted from small-amplitude responses, such as ambient vibration records or records of weak shaking from earthquakes, it can be assumed that the behavior is essentially linear. This is the case of interest in this chapter.

The data consists of N_s estimates of N_m dominant modes of vibration of a structure, obtained through reliable modal identification techniques such as MODE-ID (Beck, 1978, 1996). This data is denoted $\mathcal{D} = \{\hat{\omega}_{r,j}, \hat{\psi}_{r,j} : r = 1 \ldots N_m, j = 1 \ldots N_s\}$, where $\hat{\omega}_{r,j}$ and $\hat{\psi}_{r,j} \in \mathrm{R}^{N_o}$ are the measured modal frequency and vector of observed modeshape components, respectively, for the rth mode in the jth set of estimates.

Consider a linear structural model where the stiffness matrix K and the mass matrix M are parameterized such that they are linear functions of the vector of stiffness parameters $\theta \in \mathrm{R}^{N_\theta}$ and the vector of mass parameters $\rho \in \mathrm{R}^{N_\rho}$. That is, K and M can be expressed as

$$K(\theta) = K_0 + \sum_{k=1}^{N_\theta} K_{\theta_k} \theta_k \tag{5.1}$$

$$M(\rho) = M_0 + \sum_{k=1}^{N_\rho} M_{\rho_k} \rho_k \tag{5.2}$$

where K_{θ_k}, K_0, M_{ρ_k} and M_0 are constant matrices. The modes of vibration of a structure are given by the eigenvalues and eigenvectors of the system defined by the mass matrix M and the stiffness matrix K, and so satisfy the equation

$$\left[K(\theta) - \omega_{r,j}^2 M(\rho)\right] \phi_r = 0 \tag{5.3}$$

where $\phi_r \in \mathrm{R}^{N_d}$ is the vector of mode shape components for the rth mode shape. Thus, the vector of prediction errors $\epsilon_{r,,j}$ based on the observed modal frequencies is

given by

$$K\phi_r - \hat{\omega}_{r,j}^2 M\phi_r = \epsilon_{r,j} \tag{5.4}$$

Note that this equation requires the complete modeshapes of the system, however, typical measurements of the mode shapes do not include every degree of freedom present in the identification model. In this case, the modeshapes of the identification model are introduced as parameters to be updated (Beck et al., 2001), and are referred to as *system mode shapes* to differentiate them from the incomplete observed modeshapes, which are part of the modal data. Though the system modeshapes represent the actual underlying mode shapes of the system, they are not constrained to be eigenvectors of a structural model. The system modeshapes are related to the observed modal data and the vector of prediction errors $\epsilon_{r,,j}$ as follows:

$$\hat{\psi}_{r,j} = \Gamma\phi_r + e_{r,j} \tag{5.5}$$

where Γ is the $N_o \times N_d$ observation matrix, which relates the measured degrees of freedom to those of the identification model.

The prediction errors for Equations 5.4 and 5.5, $\epsilon_{r,j}$, and $e_{r,j}$, are modeled as independent Gaussian variables. The prediction errors for each mode are assumed to be zero mean and uncorrelated with the other modes and the prediction errors for each modal component are mutually uncorrelated and have the same variance, that is

$$\begin{aligned} \epsilon_{r,j} &\sim N(0, \sigma_r^2 I_{Nd\times N_d}) \\ e_{r,j} &\sim N(0, \delta_r^2 I_{No\times No}) \end{aligned} \tag{5.6}$$

5.2 Model Updating Using the Gibbs Sampler

Updating of the parameters of the underlying structural model, the system modeshapes, and the prediction-error variances can be efficiently performed using the Gibbs sampler approach (which will henceforth be referred to as the GS approach). The key

insight is that by correctly structuring the problem, sampling from the conditional PDFs can be performed very easily, by taking advantage of the expressions for the mean and covariance of the conditional PDF for a linear system given by Equation 4.5.

The prior PDFs for the uncertain parameters in the model class $\mathcal{M}$ are as follows:

$$p(\theta \mid \mathcal{M}) = N(\mu_\theta, \Sigma_\theta) \tag{5.7}$$

$$p(\rho \mid \mathcal{M}) = N(\mu_\rho, \Sigma_\rho) \tag{5.8}$$

$$p(\phi_r \mid \mathcal{M}) = N(\mu_{\phi_r}, \Sigma_{\phi_r}) \tag{5.9}$$

$$p(\delta_r^2 \mid \mathcal{M} = IG(\alpha_{\delta_r}, \beta_{\delta_r}) \tag{5.10}$$

$$p(\sigma_r^2 \mid \mathcal{M} = IG(\alpha_{\sigma_r}, \beta_{\sigma_r}) \tag{5.11}$$

Note that the prior PDFs, like those presented in Section 4.1.1, are conditionally conjugate, simplifying the calculations for the conditional PDFs. The procedure for using the GS algorithm for updating the linear structural model parameters is as follows:

For the first step, values for the linear structural model parameters, $\tilde{\theta}^{(0)}$ and $\tilde{\rho}^{(0)}$, and the prediction-error variances, $\tilde{\delta}_r^{(0)}$ and $\tilde{\sigma}_r^{(0)}$, $r = 1, \ldots, N_m$, are drawn from the appropriate prior PDFs. The goal is now to draw a sample from the conditional PDF for the system modeshapes, $p(\phi_r \mid \theta^{(0)}, \rho^{(0)}, \sigma_r^{(0)}, \mathcal{D})$, $r = 1, \ldots, N_m$. Note that the equations for each mode are de-coupled, so that information observed for one mode is irrelevant to all the other modes, and the calculation of the conditional PDFs can be performed independently for each mode. Also, recall from Section 2.1.3 that updating with different data sets may be performed sequentially, with the posterior PDF obtained by the first set of data used as the prior PDF for updating with the second set of data, and so on. The same approach can be applied here to update with one set of identified modal parameters at a time.

The mean and covariance matrix for the conditional PDF for the rth system modeshape updated with just the first set of data (the identified frequencies $\hat{\omega}_{r,1}^2$ and modeshape components $\hat{\psi}_{r,1}$) are calculated from Equation 4.5. The vectors X_0, X,

and Y are given by

$$X_0 = \mu_{\phi_r} \tag{5.12}$$

$$X = \phi_r \tag{5.13}$$

$$Y = \left[\begin{array}{cc} \hat{\psi}_{r,1}^T & \mathbf{0}_{1\times N_d} \end{array} \right]^T \tag{5.14}$$

where $\mathbf{0}_{1\times N_d}$ is a $1 \times N_d$ vector of zeros. The matrices A, Σ and Σ_0 are given by

$$A = \left[\begin{array}{c} \Gamma \\ K\left(\tilde{\theta}^{(0)}\right) - \hat{\omega}_{r,1}^2 M\left(\tilde{\rho}^{(0)}\right) \end{array} \right] \tag{5.15}$$

$$\Sigma = \left[\begin{array}{cc} \left(\tilde{\delta}_r^{(0)}\right)^2 I_{N_d\times N_d} & \\ & \left(\tilde{\sigma}_r^{(0)}\right)^2 I_{N_d\times N_d} \end{array} \right] \tag{5.16}$$

$$\Sigma_0 = \Sigma_{\phi_r} \tag{5.17}$$

The resulting mean and covariance matrix obtained from Equation 4.5 are denoted $\mu_{r,1}$ and $\Sigma_{r,1}$, respectively. Now, updating with the second set of data is performed in the same manner, but the vector X_0 and the matrix Σ_0 are replaced with $\mu_{r,1}$ and $\Sigma_{r,1}$, respectively. This process is repeated until the final mean and covariance matrix for the conditional PDF, μ_{r,N_s} and Σ_{r,N_s}, are obtained. Samples for the system modeshapes $\tilde{\phi}_r^{(1)}$, $r = 1, \ldots, N_m$ are then drawn from this final conditional PDF.

The next step in the GS algorithm requires drawing samples for the structural model parameters, θ and ρ, from the conditional PDF $p(\theta, \rho \mid \mathcal{D}, \tilde{\phi}_r^{(1)}, \sigma_r^{(0)})$. The approach is similar to the previous step, in that each mode and each data set may be applied separately. For updating with the first modal equation and the first set of modal data, Equation 4.5 is again used to find the mean and covariance matrix of the conditional PDF for the model parameters, with the vectors X_0, X, and Y given

by:

$$X_0 = \left[\begin{array}{cc} \mu_\theta^T & \mu_\rho^T \end{array} \right]^T \tag{5.18}$$

$$X = \left[\begin{array}{cc} \theta^T & \rho^T \end{array} \right]^T \tag{5.19}$$

$$Y = \underline{\mathbf{0}}_{N_d \times 1} \tag{5.20}$$

and the matrices A, Σ and Σ_0 given by

$$A = [a_{\theta_1} \ldots a_{\theta_{N_\theta}} a_{\rho_1} \ldots a_{\rho_{N_\rho}}] \tag{5.21}$$

$$a_{\theta_j} = K_{\theta_j} \tilde{\phi}_1^{(1)}$$

$$a_{\rho_j} = -\hat{\omega}_{1,1}^2 M_{\rho_j} \tilde{\phi}_1^{(1)}$$

$$\Sigma = (\sigma_1^{(0)})^2 I_{N_d \times N_d} \tag{5.22}$$

$$\Sigma_0 = \left[\begin{array}{cc} \Sigma_\theta & \\ & \Sigma_\rho \end{array} \right] \tag{5.23}$$

As before, the resulting mean and covariance matrix are denoted $\mu_{1,1}$ and $\Sigma_{1,1}$, and replace X_0 and Σ_0, respectively. The updating steps are repeated until the final conditional mean and covariance matrix, μ_{N_m,N_s} and $\Sigma_{Nm,Ns}$, are obtained and these are used to generate the samples $\tilde{\theta}^{(1)}$ and $\tilde{\rho}^{(1)}$.

The final step in each iteration of the Gibbs sampler is the draw for the updated prediction-error variances. As shown in Section 4.1.1, this can be done quite easily if the prior PDF for the error variances is of a form that is conditionally conjugate. The conditional PDFs for the error variances are proportional to inverse gamma distributions,

$$p(\delta_r^2 \mid \tilde{\phi}_r^{(1)}, \tilde{\theta}^{(1)}, \tilde{\rho}^{(1)}, \mathcal{D}) \propto IG\left(\alpha_\delta + \frac{N_s N_o}{2}, \beta_\delta + \sum_{i=1}^{N_s} \|\hat{\psi}_{r,i} - \Gamma \tilde{\phi}_r^{(1)}\|^2\right) \tag{5.24}$$

$$p(\sigma_r^2 \mid \tilde{\phi}_r^{(1)}, \tilde{\theta}^{(1)}, \tilde{\rho}^{(1)}, \mathcal{D}) \propto IG\left(\alpha_\sigma + \frac{N_s N_d}{2}, \beta_\sigma + \sum_{i=1}^{N_s} \left\| \left[K(\tilde{\theta}^{(1)}) - \hat{\omega}_{r,i}^2 M(\tilde{\rho}^{(1)})\right] \tilde{\phi}_r^{(1)} \right\|^2\right)$$

which are then sampled to obtain $(\delta_r^{(1)})^2$ and $(\sigma_r^{(1)})^2$. These three steps are repeated until the desired number of samples is generated.

5.2.1 Estimating the Probability of Damage

Damage, in this problem, will be defined to be a reduction in the value of any component of the vector of stiffness parameters θ below some specified threshold value when comparing the model identified with data from the possibly damaged structure, $\mathcal{D}^{pd}$ to the model identified with data from the undamaged structure, $\mathcal{D}^{ud}$. Stochastic simulation methods are used to generate N^{ud} samples for the ith component of θ, conditioned on the undamaged data, $\tilde{\theta}_{i,k}^{ud}, k = 1 \dots N^{ud}$, and N^{pd} samples conditioned on the possibly damaged data, $\tilde{\theta}_{i,k}^{pd}, k = 1 \dots N^{pd}$. The probability that the damage fraction is at least d in the parameter component θ_i may be approximated using M comparisons between the values of samples of the parameter θ_i generated from the two sets of data, as follows:

$$P(\theta_i^{pd} < (1-d)\theta_i^{ud} \mid \mathcal{D}^{ud}, \mathcal{D}^{pd}) \approx \frac{1}{M} \sum_{j=1}^{M} I[\tilde{\theta}_{i,n_j}^{pd} < (1-d)\tilde{\theta}_{i,m_j}^{ud}] \tag{5.25}$$

where $I[.]$ is the indicator function, which is unity when the condition is satisfied and zero otherwise and the samples $\tilde{\theta}_{i,m_j}^{ud}$ and $\tilde{\theta}_{i,n_j}^{pd}$ are chosen randomly from the available samples conditioned on the undamged and possibly damaged data, respectively, for the jth comparison. In this study, an equal number of samples is generated for the undamaged and possibly damaged cases ($N^{ud} = N^{pd} = N$), so a full comparison of all of the samples would require $M = N^2$ evaluations. It was found in this work, however, that the estimate of the probability of damage converged adequately to the final value after randomly selecting $M = N$ sample pairs for comparison.

5.3 2-DOF Shear-Building Example

To examine the performance of the Gibbs sampler algorithm, studies are performed using simulated data from a simple 2-DOF linear shear building. The lumped masses of each floor are equal, and the ratio of inter-story stiffness to mass is 2,000 : 1 for each floor. A Rayleigh viscous-damping matrix is chosen to give a damping ratio of 2% to both modes of vibration.

Data is generated from the undamaged model and two models with simulated damage. Damage pattern DP 1 has a 30% reduction in the inter-story stiffness of the first story, while DP 2 has a 30% reduction in the second-story stiffness. The mass and damping matrices are unchanged in the damaged cases.

The three simulated structures are subjected to 110 seconds of white-noise ambient excitation at each degree of freedom, and the time histories of the acceleration responses are calculated. Gaussian-distributed noise is added to each measurement, with the standard deviation of the noise equal to 10% of the RMS (root-mean-square) of the response record that it is added to.

Modal properties for each structure were estimated from the acceleration records using the program MODE-ID (Beck, 1996). The first 10 seconds of each record were discarded to eliminate possible transients, and the remaining 100 seconds were divided into $N_s = 10$ segments, with modal properties identified for each segment. The results of the identification are summarized in Table 5.1, which shows the mean and standard deviation of the modal frequencies and ratios of the modeshape components identified for each the 10 segments.

To study the effect that missing information has on the performance of the GS method, two cases are considered. In the first case, a data set $\mathcal{D}_{full}$ contains modal frequencies and modeshape components for both modes, representing the case where sensors are available on both floors and both modes can be clearly identified. In the second case, a data set $\mathcal{D}_{part}$ contains only the first modal frequency, representing the case where only a single sensor is available and only the first mode is identifiable.

Damage Pattern	Frequency (Hz)		Modeshape ratio $\phi_{r,2}/\phi_{r,1}$	
	1st Mode	2nd Mode	1st Mode	2nd Mode
Undamaged	4.43 ±0.13	11.59 ±0.35	1.58 ±0.07	-0.44 ±0.12
DP 1	3.87 ±0.12	11.14 ±0.28	1.40 ±0.03	-0.42 ±0.12
DP 2	4.17 ±0.12	10.42 ±0.26	1.92 ±0.09	-0.31 ±0.14

Table 5.1: Identified modal frequencies and modeshape component ratios for the 2-DOF simulated shear building.

5.3.1 Damage Identification

A 2-DOF linear shear-building model is used for identification. The masses are controlled by the parameters (ρ_1, ρ_2), and the inter-story stiffnesses by (θ_1, θ_2), as follows:

$$M = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} K = \begin{bmatrix} 2000\theta_1 + 2000\theta_2 & -2000\theta_2 \\ -2000\theta_2 & 2000\theta_2 \end{bmatrix} \tag{5.26}$$

Additionally, for each mode present in the data, a system modeshape vector ϕ_r and prediction-error variance σ_r are identified.

The prior PDFs of (ρ_1, ρ_2) are chosen to be independent Gaussian PDFs, each with a mean of 1 and a coefficient of variation (c.o.v., the standard deviation expressed as a fraction of the mean) of 10%. The prior for each modeshape component was taken to be an independent Gaussian PDF with a mean of zero and large c.o.v., giving a relatively flat prior. Inverse gamma distributions were taken for the prior PDFs of the prediction-error variances, with shape parameters $\alpha_{\delta_r} = \alpha_{\sigma_r} = 3$ and scaling parameters β_{δ_r} and β_{σ_r} chosen to scale the prior PDFs to the variances of the measured data for each mode.

Using this parameterization, setting the mass and stiffness parameters equal to zero would result in an unwanted solution to Equation 5.4. For this example, the chosen priors are "far" enough from this trivial solution that this is not a concern. However, if the non-trivial solution were located closer to zero, it would cause problems. This could be addressed by assuming a lognormal distribution for the model parameters and treating the logarithm of the model parameters as the uncertain

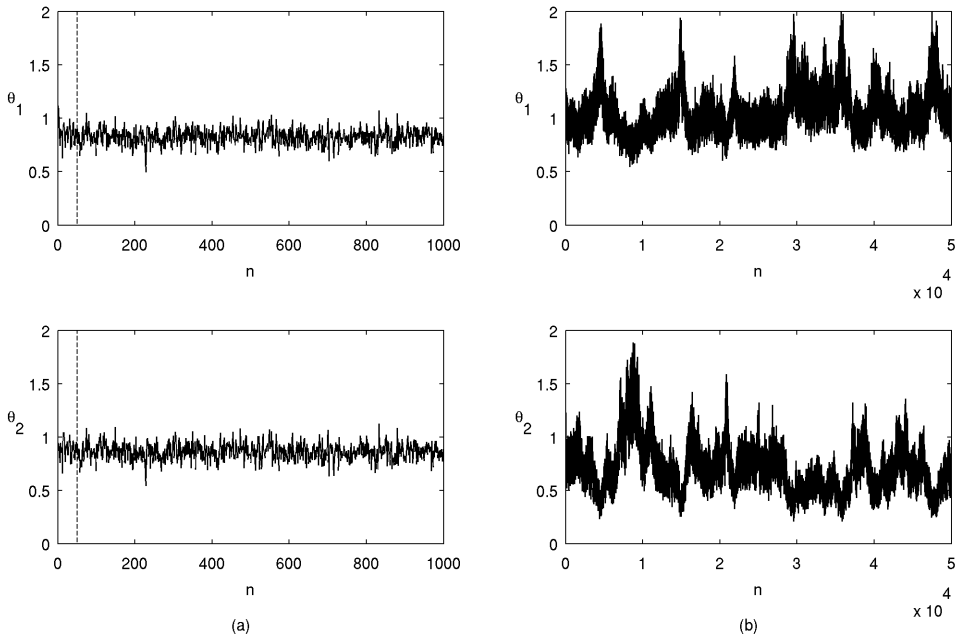


Figure 5.1: Markov chain samples for the stiffness parameters generated using the data from the undamaged 2-DOF model updating with (a) $\mathcal{D}_{full}$ (globally identifiable) and (b) $\mathcal{D}_{part}$ (unidentifiable). The index n denotes the step number of the Markov chain. The dashed lines in (a) indicate the end of the burn-in period at $n = 50$ (not visible for the unidentifiable cases because of scale).

parameters.

The GS algorithm is used to generate five sets of samples from the posterior PDF of the model parameters for each of the data sets $\mathcal{D}_{full}$ and $\mathcal{D}_{part}$ in the undamaged case and the two damaged cases DP 1 and DP 2. For $\mathcal{D}_{full}$, 1000 samples are used in each chain, while 50,000 samples per chain are generated for $\mathcal{D}_{part}$, for reasons that will be discussed later. Figure 5.1(a) shows the Markov chain samples for the stiffness parameters generated during run one using the data $\mathcal{D}_{full}$ with the undamaged model, and shows the end of the burn-in period, chosen by inspection to be 50 samples. Figure 5.1(b) shows samples generated during one run with the undamaged model using $\mathcal{D}_{part}$. The same burn-in period of 50 samples is chosen, but is not visible in Figure 5.1(b) because of scale.

The identifiability of each case is immediately clear if the samples for the stiffness parameters are plotted in the (θ_1, θ_2) space. Figure 5.2(a) shows that the samples of

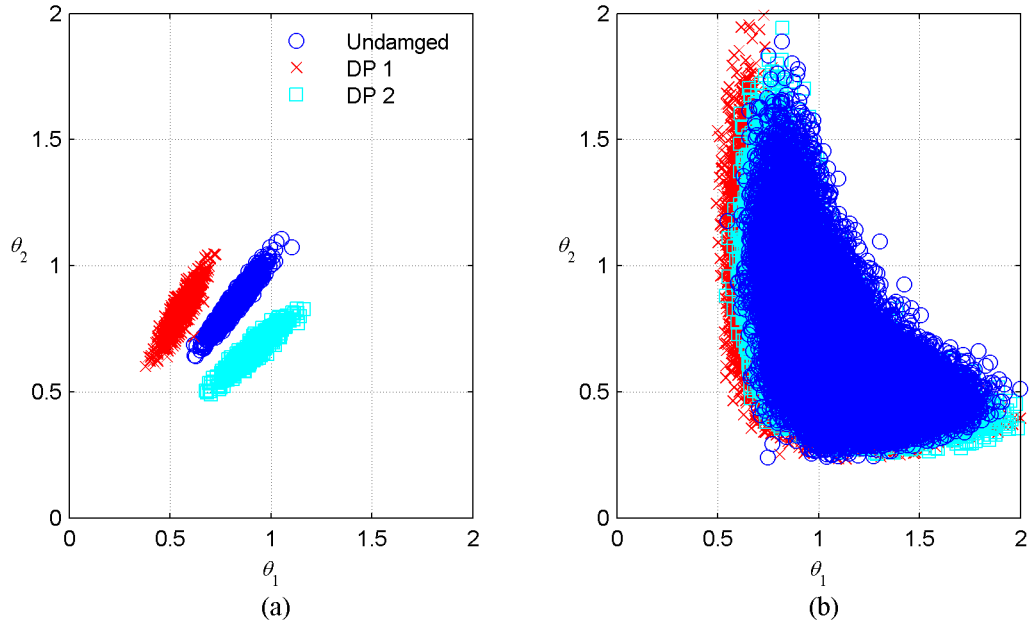


Figure 5.2: Stiffness parameter samples for the 2-DOF model plotted in the (θ_1, θ_2) space for all damage cases (undamged, DP 1, DP 2) for (a) $\mathcal{D}_{full}$ and (b) $\mathcal{D}_{part}$.

the stiffness parameters generated using $\mathcal{D}_{full}$ are concentrated into a relatively small area, indicating global identifiability. The apparent interaction between the stiffness parameters is due to the fact that the mass parameters are also uncertain. On the other hand, the samples generated using $\mathcal{D}_{part}$, shown in Figure 5.2(b), are spread out, indicating that there is insufficient data to determine the stiffness parameters and so giving an unidentifiable case. Damage probability curves are calculated for each case using Equation 5.25. Figures 5.3 and 5.4 show the damage probability curves for the cases with 1st-story damage and 2nd-story damage, respectively. A useful summary of the information contained in a damage probability curve is the *median damage fraction*, which is given by the value that has a 50% probability of exceedance (the value on the x-axis of Figures 5.3 and 5.4 that corresponds with the intersection of the damage probability curve with the line $y = 0.5$). It is immediately apparent that the samples generated using $\mathcal{D}_{full}$ can be used to determine the presence and location of damage. The damage curve generated for DP 1 in Figure 5.3 has a median damage fraction of 0.33, and the curve for DP 2 in Figure 5.4 has a median damage fraction of

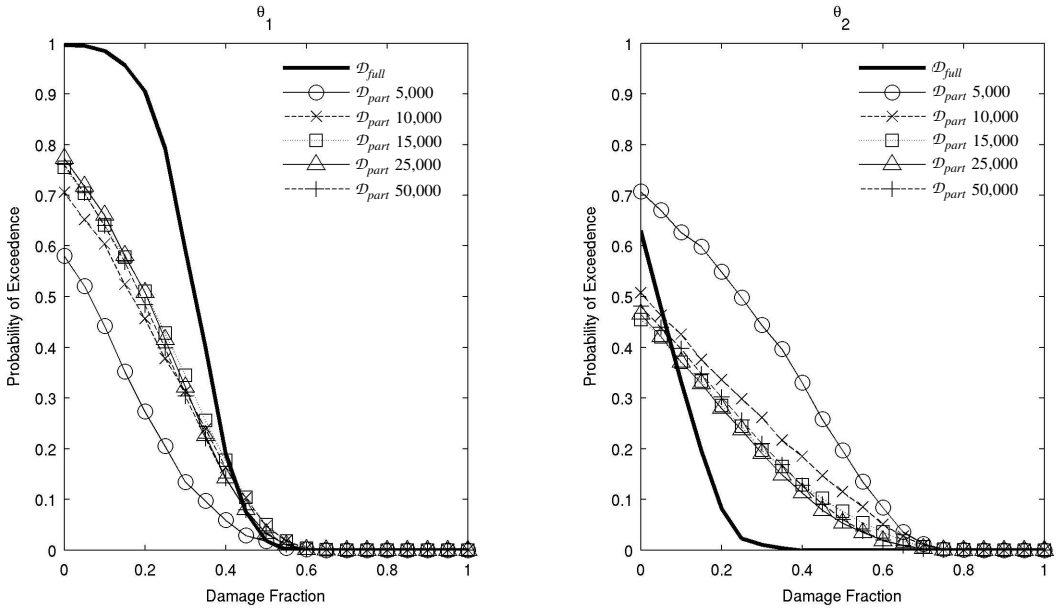


Figure 5.3: Estimated damage probability curves for DP 1.

0.22, with both values being roughly the same as the actual damage fraction, which was 0.30 in both cases.

The cases where $\mathcal{D}_{part}$ was used to generate the samples are more complicated. It's clear from Figures 5.3 and 5.4 that a larger number of samples is required for the results of the damage curve to converge. In the case of DP 1, 15,000 samples seems to be adequate, with a median damage fraction of 0.20. However, for DP 2, convegence is slower and damage detection is difficult: for 15,000 samples, the median damage fractions are close to being equal; 0.04 for the first story and 0.05 for the second story. When 25,000 samples are considered, these damage fractions are 0.02 for the first story and 0.09 for the second story. When all 50,000 are used, the mean damage fraction for the first story is 0.06 and the second story is identified as undamaged.

To examine the variability of the estimated median damage fraction if only one run of the GS algorithm is performed for the undamaged and possibly damaged case, damage probabilty curves were calculated for each of the 25 possible combinations of undamaged and possibly damaged runs for each damage pattern (5 runs of each are available). The mean and standard deviation of the median damage fractions

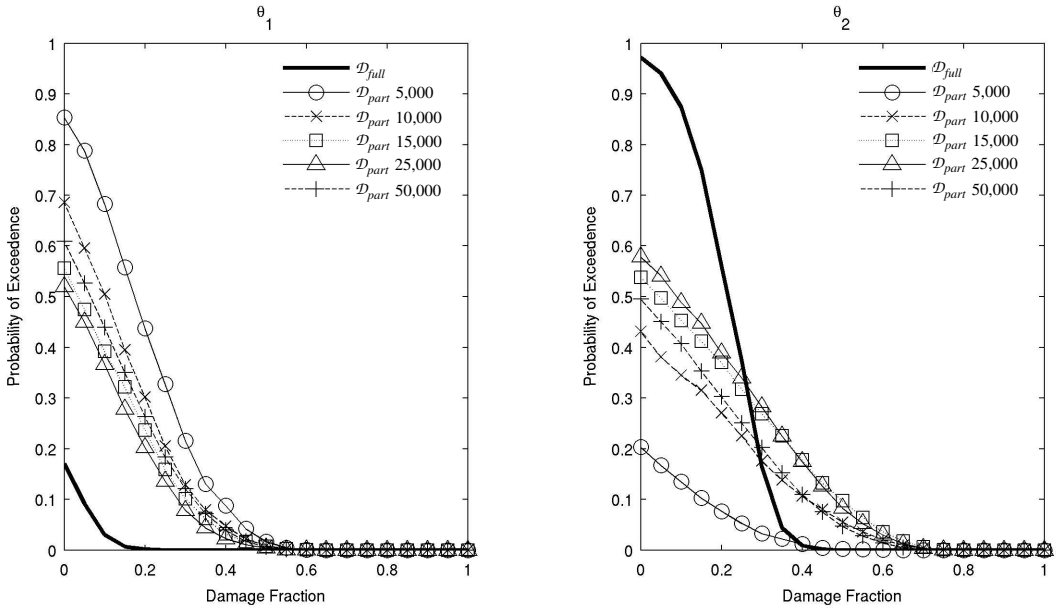


Figure 5.4: Estimated damage probability curves for DP 2.

obtained for these sets of damage probability curves are given in Table 5.2.

These results show that knowing only a single modal frequency gives insufficient information for damage assessment. Insight into the results for the cases using $\mathcal{D}_{part}$ is given by considering the geometry of the detection problem in the (θ_1, θ_2)-space. Figure 5.5 shows curves that represent combinations of θ_1 and θ_2 in the identification model which will (when the mass parameters are fixed to unity) give the correct identified frequency of the structure in each of the damage states considered. Clearly, the values of (θ_1, θ_2) along the curves for ω_1 in Figure 5.5 are nearly indistinguishable on the basis of $\mathcal{D}_{part}$ alone. The results in Table 5.2, however, show that there is some indication of the presence of damage, particularly for DP 1, which, as seen in Figure 5.5 has a curve for constant ω_1 that is "farther" away from the curve for the undamaged system than the curve for constant ω_1 for DP 2.

5.3.2 Locally Identifiable Models

If we consider the curves for constant values of ω_1 and ω_2 in Figure 5.5, we see that there should be two points in the (θ_1, θ_2)-space where the curves intersect. The 2-

Data	# Samples	Damage Pattern	Median Damage Fraction	
			θ_1	θ_2
$\mathcal{D}_{full}$	1000	DP 1	0.323 ± 0.005	0.040 ± 0.008
		DP 2	0.000 ± 0.000	0.221 ± 0.006
$\mathcal{D}_{part}$	5000	DP 1	0.175 ± 0.094	0.081 ± 0.105
		DP 2	0.145 ± 0.081	0.013 ± 0.044
$\mathcal{D}_{part}$	10000	DP 1	0.181 ± 0.070	0.056 ± 0.085
		DP 2	0.112 ± 0.074	0.033 ± 0.068
$\mathcal{D}_{part}$	15000	DP 1	0.204 ± 0.061	0.033 ± 0.056
		DP 2	0.108 ± 0.063	0.020 ± 0.043
$\mathcal{D}_{part}$	25000	DP 1	0.218 ± 0.048	0.016 ± 0.038
		DP 2	0.106 ± 0.050	0.014 ± 0.033
$\mathcal{D}_{part}$	50000	DP 1	0.235 ± 0.035	0.003 ± 0.014
		DP 2	0.113 ± 0.041	0.004 ± 0.011

Table 5.2: Estimated median damage fractions for the 2-DOF examples.

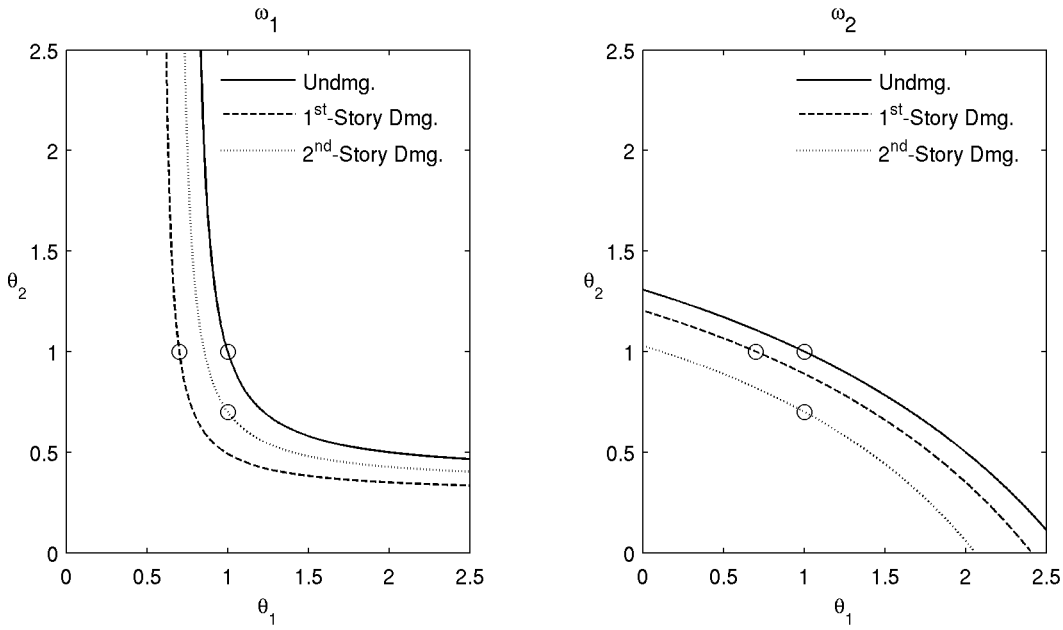


Figure 5.5: Curves in the (θ_1, θ_2)-space along which a given natural frequency (ω_1 or ω_2) is held constant. The circles indicate the actual state of the structure.

DOF shear-building models defined by these points of intersection have the same natural frequencies, but different modeshapes. Therefore, if only the first two modal frequencies are available, the system should in theory be locally identifiable, with disconnected high-probability regions around each of the points of intersection. However, successfully representing this posterior PDF using sampling methods is difficult. The GS algorithm generates samples that are relatively close together, so the samples would tend to be clustered in one of the high-probabilty regions, with little chance that the sampler would "wander" far enough to discover and populate the second high-probability region. Hybrid algorithms that add "jumps" into the sampler may be effective, but detract from the qualities that make the GS algorithm so attractive: the ability to operate in relatively high dimensions efficiently. It can be argued that locally identifiable cases are probably uncommon in practice since they exist on the "boundary" between identifiable and unidentifiable cases. However, studies were made to see if adjustments to the GS approach can be made to detect multiple separated regions of high probability.

Consider the case where only one sensor is available, but both modes have been identified; that is, both modal frequencies ω_1 and ω_2 are available, but not the modeshape information. The curves in Figure 5.5 indicate that for the undamaged model, there are two points in the (θ_1, θ_2)-space that will result in models with identical natural frequencies, (1, 1) and (2, 0.5). Figure 5.6 shows the histograms of the samples of the stiffness parameters from five simulation runs with 2000 samples each that are generated using ω_1 and ω_2 data from the undamaged model. Only the solution at (1, 1) is apparent. Recall, though, that the prior distribution is centered at (1, 1), and is therefore biasing the results of the sampler to that solution.

For comparison, a similar set of samples was generated from the same data using prior PDFs for θ_1 and θ_2 with a mean of 2 and standard deviations of 1 and 0.5, respectively. To avoid converging to the trivial solution at (0,0), the coefficient of variation on the mass parameters is reduced from 10% to 5%. Histograms of the samples of the stiffness parameters are shown in Figure 5.7. The presence of two distinct regions of high probability content are clearly visible, though the relative

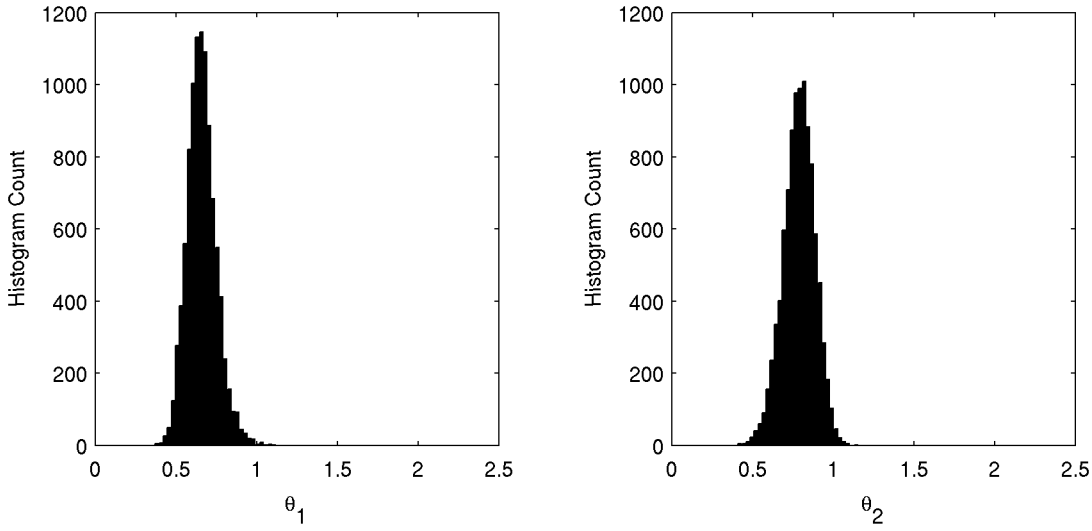


Figure 5.6: Samples generated using locally identifiable data from the undamaged model.

distribution of samples between regions does not exactly match the expected value.

While this example demonstrates that the GS approach is theoretically capable of generating samples from multi-modal posterior PDFs, in practice it is not very robust. The ability of the GS algorithm to populate multiple high-probability regions is very sensitive to the choice of prior PDFs, to the point where the user is almost required to know beforehand the approximate location of each region in order to choose the appropriate prior distributions. More general techniques, such as the TMCMC algorithm, are better able to deal with multi-modal data, but even for these methods, limitations and complications can arise, especially in higher dimensions.

5.3.3 Model-Class Selection

To examine the performance of the GS algorithm, a candidate pool of two model classes is considered. The first model class, $\mathcal{M}_1$, is the four-parameter 2-DOF shear-building model considered previously. The second model class, $\mathcal{M}_2$, is nearly identical; however, the inter-story stiffnesses are constrained to be equal, that is, $\theta_1 = \theta_2 = \theta$. Note that in the case where only data from a single sensor is available ($\mathcal{M}_1$ is unidentifiable), model class $\mathcal{M}_2$ is still globally identifiable, as can be seen by

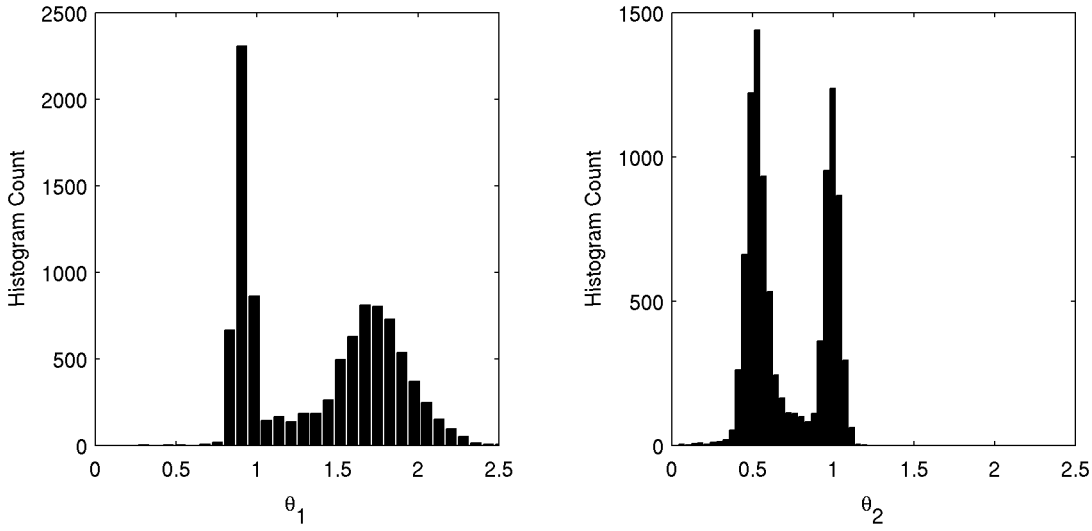


Figure 5.7: Samples generated using locally identifiable data from the undamaged model with a shifted prior PDF.

examining Figure 5.5.

The evidence for each model class is estimated from samples generated with the GS algorithm as outlined in Equations 3.3 through 3.5 in Section 3.3. The results are shown in Table 5.3.

For cases using $\mathcal{D}_{full}$, model class $\mathcal{M}_2$ dominates for selections with data generated from the undamaged structure. Since the inter-story stiffnesses actually are equal in this case, the two model classes are expected to be able to fit the data equally well, so the preference for the simpler model class in the evidence is reasonable. In contrast, for the cases involving damage, the simpler model class $\mathcal{M}_2$ cannot match the data as well as the more flexible model class $\mathcal{M}_1$.

When the data set $\mathcal{D}_{part}$ is used to generate the samples, the results for the damaged cases are very similar, though model class $\mathcal{M}_1$ is not as strongly preferred as when $\mathcal{D}_{full}$ is used, since the effect of the data-fitting performance of the candidate model classes is somewhat reduced by the smaller amount of available data. However, for the undamaged case, the more complicated model class $\mathcal{M}_1$ is preferred to the simpler $\mathcal{M}_2$, which seems counter-intuitive when compared to the results obtained using $\mathcal{D}_{full}$.

Data	Dmg. State	Probability	
		$\mathcal{M}_1$	$\mathcal{M}_2$
$\mathcal{D}_{full}$	Undmg.	0.235	0.765
	DP 1	1.000	0.000
	DP 2	1.000	0.000
$\mathcal{D}_{part}$	Undmg.	0.875	0.125
	DP 1	0.720	0.280
	DP 2	0.808	0.192

Table 5.3: Model-class selection results for 2-DOF system

The result for the undamaged case using $\mathcal{D}_{part}$ may be understood by examining the information-theoretic interpretation of the evidence given by Equation 2.16. In this case, the data fit is roughly equal for each model, but examination of the information gain for each model class shows that for the chosen prior PDFs over the stiffness parameters, less information is extracted for model class $\mathcal{M}_1$. This is because the samples for $\mathcal{M}_1$ are generated around a broad, curved region in the (θ_1, θ_2)-space similar in geometry to the broad prior PDF, as shown in Figure 5.8(a), while model class $\mathcal{M}_2$ is globally identifiable, with samples concentrated into a relatively small regiong, compared to the prior PDF, as shown in Figure 5.9. Consequently, the posterior distribution for $\mathcal{M}_1$ is "closer" to the chosen prior than the posterior distribution for $\mathcal{M}_2$. The above result implies that a different prior PDF over the stiffness parameters could significantly alter the results of Bayesian model-class selection. To verify this, updating is performed for all cases using a prior PDF centered on $(2, 2)$ in the (θ_1, θ_2)-space. While the resulting samples are essentially indistinguishable from those generated during the original analysis (after discarding the burn-in samples), the results for model-class selection, shown in Table 5.4, are considerably different for some of the cases.

For the cases using $\mathcal{D}_{full}$, there is a stronger preference for model class $\mathcal{M}_2$ in the undamaged case, while the results for DP 1 and DP 2 are essentially the same as before. Model class $\mathcal{M}_1$ is still preferred in the undamaged case using $\mathcal{D}_{part}$, though the probability is slightly lower, but for the probabilities for DP 1 and DP 2 have changed significantly. Model class $\mathcal{M}_2$ is now preferred for DP 1 and the probability of model class $\mathcal{M}_1$ in DP 2 drops from 81% to 61%. By considering Figure 5.5,

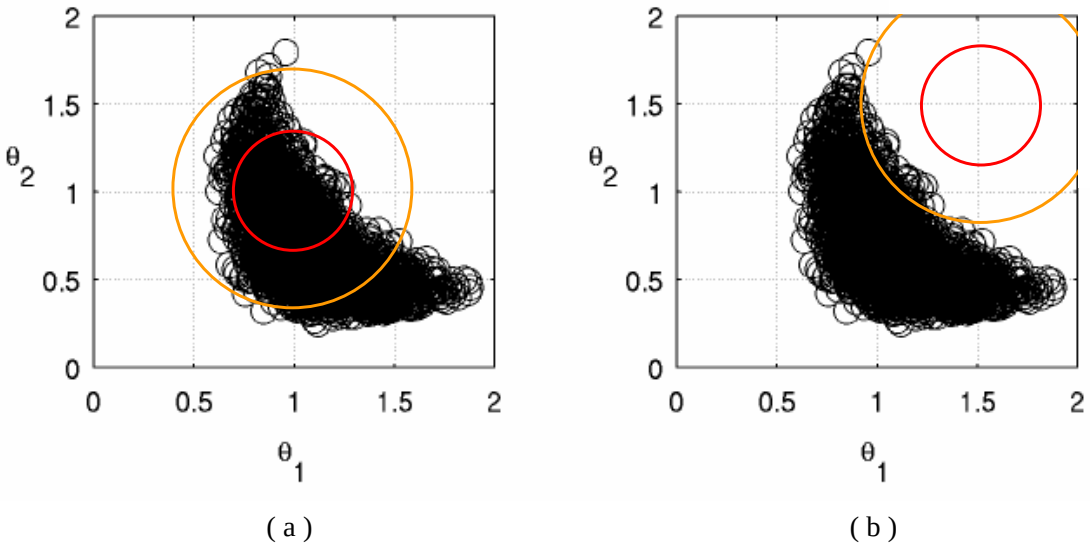


Figure 5.8: Comparison of the posterior samples to the prior PDF for (a) the original prior centered at (1,1) and (b) the shifted prior, centered at (1.5,1.5). The circles indicate points one and two standard deviations from the mean of the prior PDF.

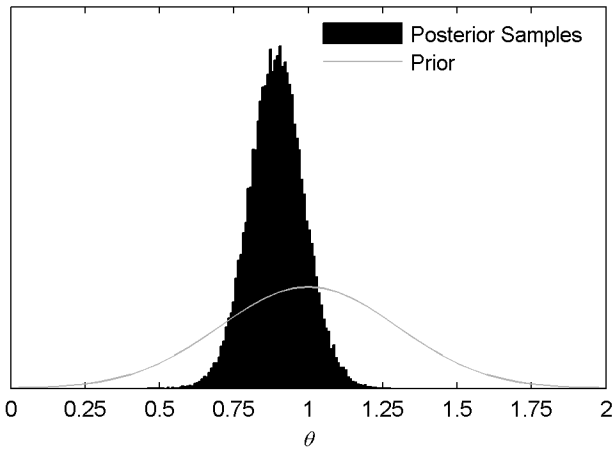


Figure 5.9: Comparison of the prior PDF and the marginal posterior PDF (calculated from samples) for the stiffness parameter for model class $\mathcal{M}_2$.

Data	Dmg. State	Probability	
		$\mathcal{M}_1$	$\mathcal{M}_2$
$\mathcal{D}_{full}$	Undmg.	0.028	0.972
	DP 1	1.000	0.000
	DP 2	1.000	0.000
$\mathcal{D}_{part}$	Undmg.	0.832	0.168
	DP 1	0.383	0.617
	DP 2	0.609	0.391

Table 5.4: Model-class selection results for 2-DOF system with shifted prior PDFs.

Data	# Samples	Dmg. State	Median Damage Fraction
$\mathcal{D}_{full}$	1000	DP 1	0.188 ± 0.004
		DP 2	0.253 ± 0.005
$\mathcal{D}_{part}$	1000	DP 1	0.148 ± 0.008
		DP 2	0.063 ± 0.010

Table 5.5: Estimated median damage fractions for model class $\mathcal{M}_2$.

it becomes apparent that the first natural frequency is more sensitive to first-story damage than second-story damage. Therefore, the identified stiffnesses for DP 1 are lower than for DP 2, and consequently "further" away from the prior PDF over the stifness parameters, given the simpler model class $\mathcal{M}_2$ the advantage.

It should be noted that this model-class selection criteria determines the most probable model class for matching the identified modal information, rather than the usefulness for damage detection. Table 5.5 gives estimates for median damage fractions (and the standard deviations of those estimates) based on samples generated for model class $\mathcal{M}_2$, and calculated in the same manner as the values given in Table 5.2. Compared to the results for model class $\mathcal{M}_1$ shown in Table 5.2, model class $\mathcal{M}_2$ seems to be less sensitive to damage when $\mathcal{D}_{part}$ is used for updating.

When choosing model classes for use in SHM applications, it may be more informative to use a selection criteria based on the ability of the model class to correctly detect and localize damage. In this case, the data $\mathcal{D}$ would consist of sets of ambient structural vibrations from a variety of damaged and undamaged models, where the damage state is known and is also part of $\mathcal{D}$. The likelihood function would then measure the ability of the model classes to correctly predict the presence, location and severity of damage. This approach to model-class selection would be very similar to the "training" of neural networks (Lam et al., 2006).

5.4 IASC-ASCE Structural Health Monitoring Simulated Benchmark Structure

The second case examined is the application of the GS approach to the connection-damaged cases in the IASC-ASCE phase II simulated structural health monitoring benchmark problems (Bernal et al., 2002; Ching et al., 2004). Data is generated by a 120-DOF finite-element structural model of a four-story, two-bay by two-bay structure, referred to as the benchmark model. The data corresponds to accelerations measured at the four faces of each of the four floors above the base level; that is, two sensors point in each of the x and y directions at each floor.

The connection-damage cases in the Phase II Benchmark problem involve the detection and assessment of simulated beam-column connection damage with different severeties and different locations in the benchmark model. The damage is simulated by reducing the stiffness of the rotational springs modelling the beam-column connections. Two of the damage patterns presented in the benchmark case are considered: DP 1 features complete loss of rotational stiffness at six first-floor beam-column connections and four second-floor beam-column connections, as shown in Figure 5.10(a), while DP 2 features complete loss of rotational stiffness at four first-story beam-column connections, as shown in Figure 5.10(b). The connection-damage cases are the most challenging in the simulated benchmark study, as the rotational degrees of freedom are not directly observed.

The data consist of 210 seconds of simulated ambient response generated for each model. Similar to the procedure used in the 2-DOF example, the first ten seconds were discarded and the remainder of the records were partitioned into ten 20-second segments. Modal parameters were identified for each segment using MODE-ID.

Eight modes were identified for each damage case, four each in the weak (x) and strong (y) directions (as labeled in Figure 5.10). Table 5.6 shows the identified modal frequencies for each case. Note that the identified frequencies of the modes of vibration in the weak direction are essentially unchanged by damage patterns DP 1 and DP 2, as expected, since the damage corresponds to reductions of the rotational stiff-

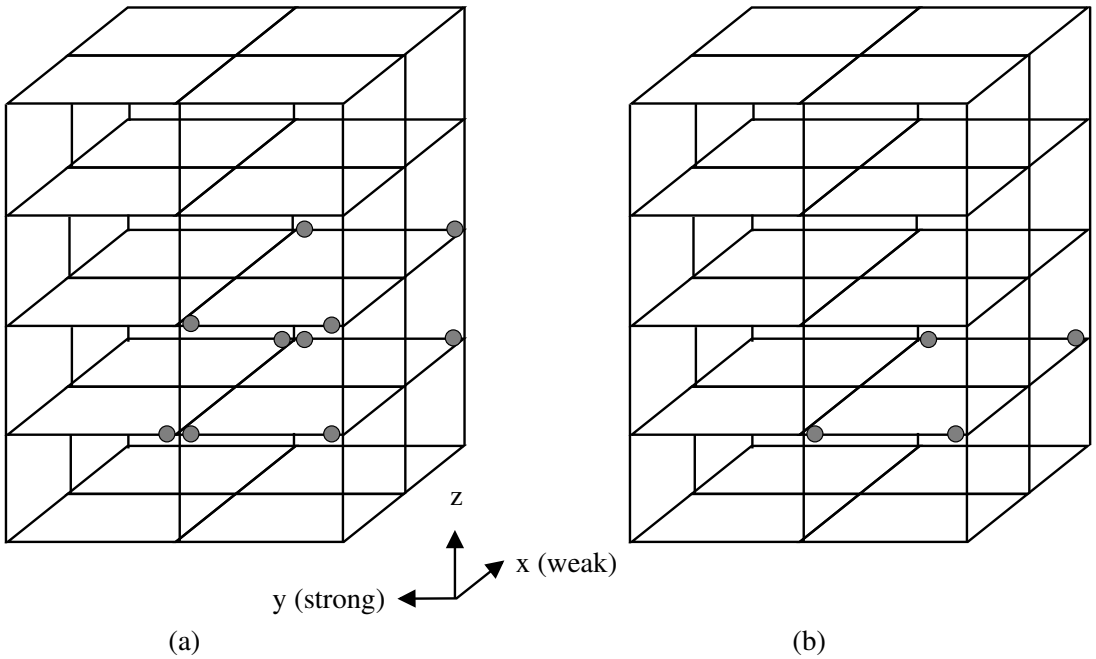


Figure 5.10: Diagram of benchmark structure showing damage locations for (a) DP 1 and (b) DP 2. The circles indicate connections with reduced rotational stiffness. Taken from Ching et al. (2006).

nesses, affecting only the lateral stiffness in the strong direction. Figure 5.11 shows the identified mode shapes. Only the translational components are directly identifiable from the sensor data. Notice that the damage has virtually no effect on these mode shape components, indicating that essentially only the decreased frequencies for modes S1 through S4 give information about the damage.

A 3-D 36-DOF identification model is used. The floors are assumed to be rigid for in-plane deformation in the x-y plane but rotation along the x and y axes is allowed.

Damage Pattern	Frequency (Hz)							
	W1	S1	W2	S2	W3	S3	W4	S4
Undamaged	3.19 ±0.06	3.98 ±0.01	9.79 ±0.05	13.41 ±0.04	16.66 ±0.19	25.15 ±0.09	23.72 ±0.09	39.28 ±0.08
DP 1	3.20 ±0.08	3.42 ±0.03	9.78 ±0.06	12.91 ±0.08	16.69 ±0.07	24.68 ±0.10	23.72 ±0.09	39.11 ±0.12
DP 2	3.19 ±0.07	3.79 ±0.02	9.79 ±0.05	13.13 ±0.08	16.72 ±0.08	25.15 ±0.08	23.72 ±0.09	39.17 ±0.12

Table 5.6: Modal frequencies for the benchmark structure. “W” and “S” denote the “weak” and “strong” directions, respectively.

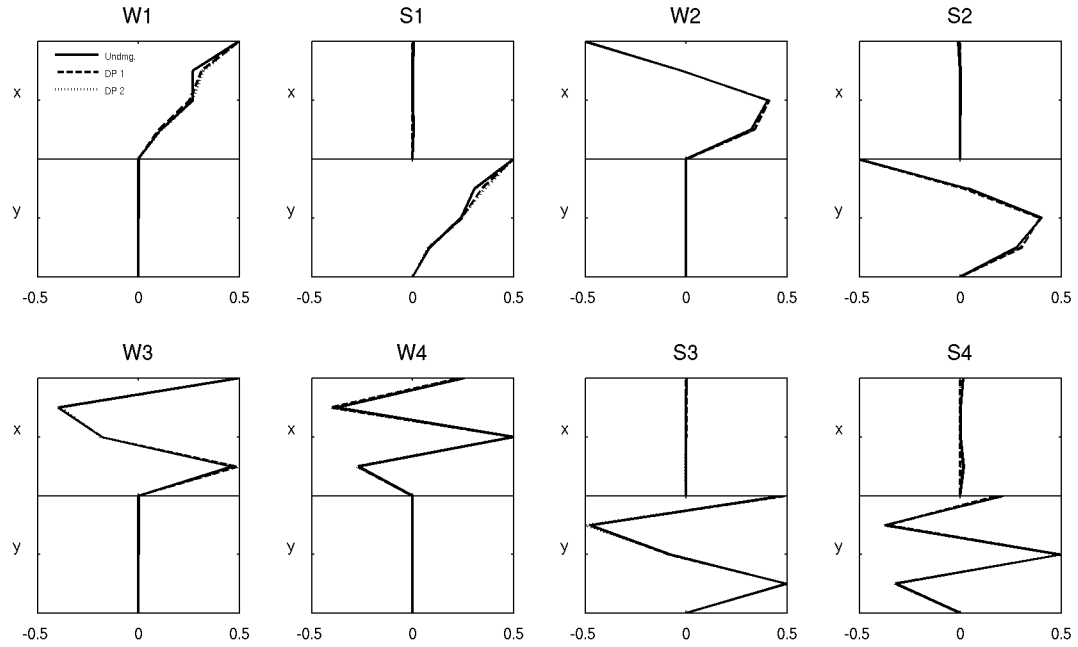


Figure 5.11: Mean identified mode shapes of the benchmark structure.

A diagram of the degrees of freedom for each floor is shown in Figure 5.12. Three DOF are used to describe the motion of the ith floor, the translations x_i and y_i and the rotation ϕ_i about the z axis. The remaing six degrees of freedom for each floor describe the common local rotations about the x and y axes for the degrees of freedom in each row. For example, in Figure 5.12, the rotation about the x axis at points 1, 2, and 3 is given by $\gamma_{i,1}$ and the rotation about the y axis at points 3, 6, and 9 is given by $\delta_{i,3}$.

Two parameters are used for the rotational stiffness of each floor: One controlling the rotational stiffness of all beam-column connections about the x axis and the other controlling the rotational stiffness of the connections about the y axis. Four parameters are used to control the column stiffness, each one specifying the stiffness of the columns in one story. Four parameters are used to specify the mass of each floor, which represents the mass of the floor slab plus the apportioned lumped masses of the connected columns.

The prior PDF for each of the eight rotational stiffness parameters are taken to be an independent Gaussian distribution, with a mean of 1 and a c.o.v. of 20%. The

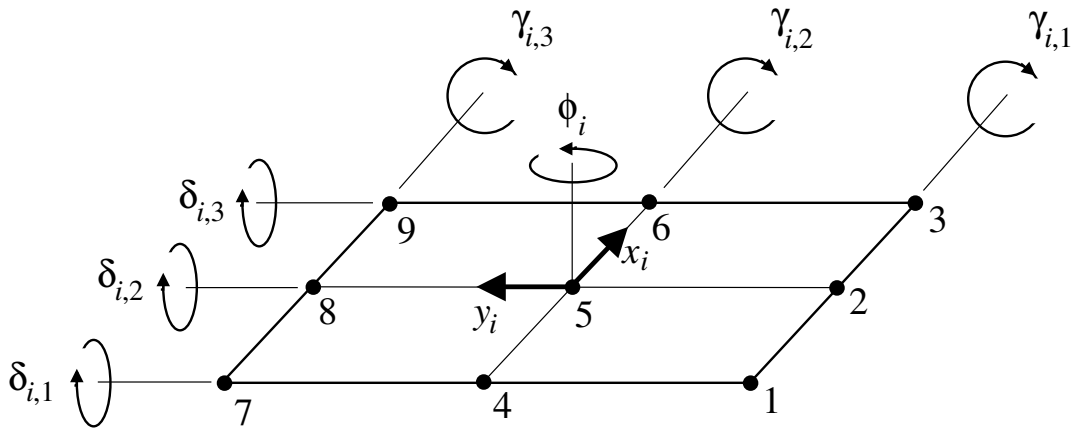


Figure 5.12: Degrees of freedom for the ith floor in the identification model.

columns dominate the global stiffness of the structure and also provide rotational stiffness, so slight errors in the identified column stiffness can significantly influence the identification of rotational stiffnesses. To reliably detect rotational stiffness damage, the prior distributions on the column stiffnesses must be relatively constrained. The prior PDF for each column stiffness parameter is taken to be an independent Gaussian distribution with a mean of 1 and a c.o.v. of 2%. Given that design of moment-frame buildings uses a "strong column-weak beam" principle, it seems reasonable to allow variations in the rotational stiffnesses, which are tied to the column-beam connections that should be damaged, to be larger than variations in the column stiffnesses.

It is assumed that prior information about the mass of the building is very reliable, so that the prior PDF has a small variance. In this case, the prior for each mass parameter is an independent Gaussian distribution with a mean of 1 and a c.o.v. of 1%. This choice of prior ensures that the mass parameters will be essentially fixed at the nominal mean values.

Following the GS approach, Markov chain samples of the 16 structural parameters (8 rotational stiffnesses, 4 column stiffnesses, 4 masses), the system modeshapes and the prediction-error variances of each mode are obtained. Since each of the 8 modes has 36 degrees of freedom, the total number of uncertain parameters is 312. Five parallel chains of 5000 samples were simulated. Figure 5.13 shows the samples of the rotational stiffness parameters for the undamaged case. A burn-in period of 1000

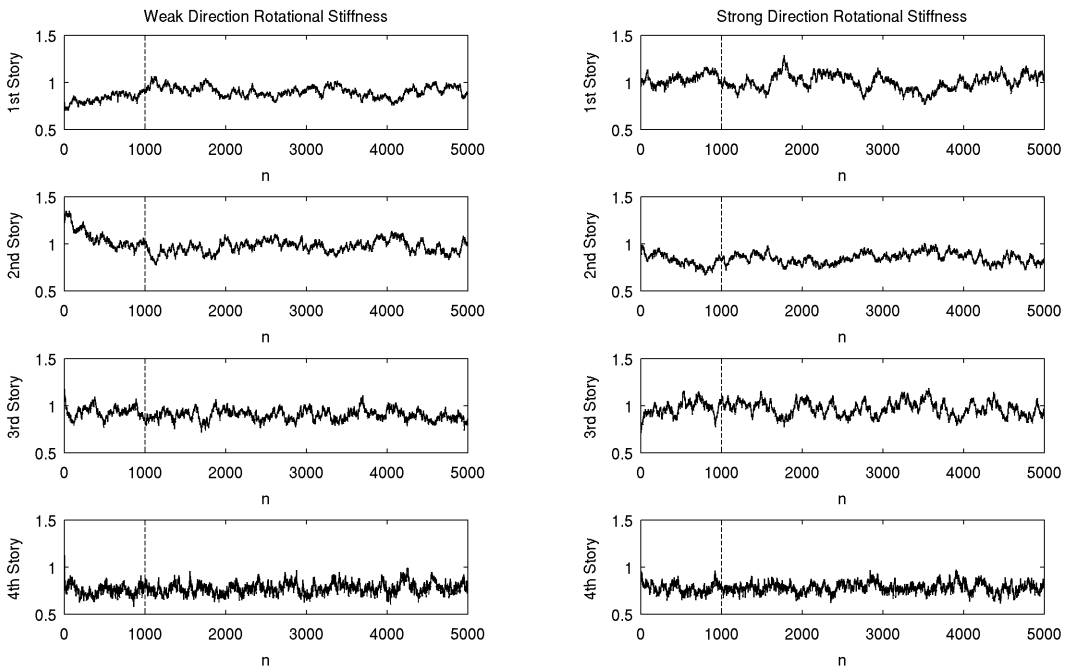


Figure 5.13: Markov chain samples for the rotational stiffness parameters generated using the data from the undamaged benchmark model. The index n denotes the step number of the Markov chain. The dashed lines indicate the end of the burn-in period at $n = 1000$.

samples was chosen after visual inspection.

The sample means of the translational components of the updated system modeshapes components are plotted for all damage states in Figure 5.14. As expected from the identified modeshapes, the damage has no discernable effect. Figure 5.15 shows the samples means of the average rotational components of the updated system modeshapes, which are calculated for the ith floor by averaging the components associated with the rotations $\gamma_{i,j}, j = 1, 2, 3$ (see Figure 5.12) for the weak (x) direction and $\delta_{i,j}, j = 1, 2, 3$ in the strong (y) direction. The results indicate that the damage alters the rotational components of the system modeshapes, particularly for the weak direction rotations for the first strong-direction modeshape. This is reasonable, since both DP 1 and DP 2 feature damage at strong-direction beam-column connections, which would be reflected in increased rotations about the perpendicular (weak) direction. Insight into the overall performance of the GS approach can be gained by examining the posterior PDFs of the rotational stiffness parameters. Figure 5.16 shows kernel

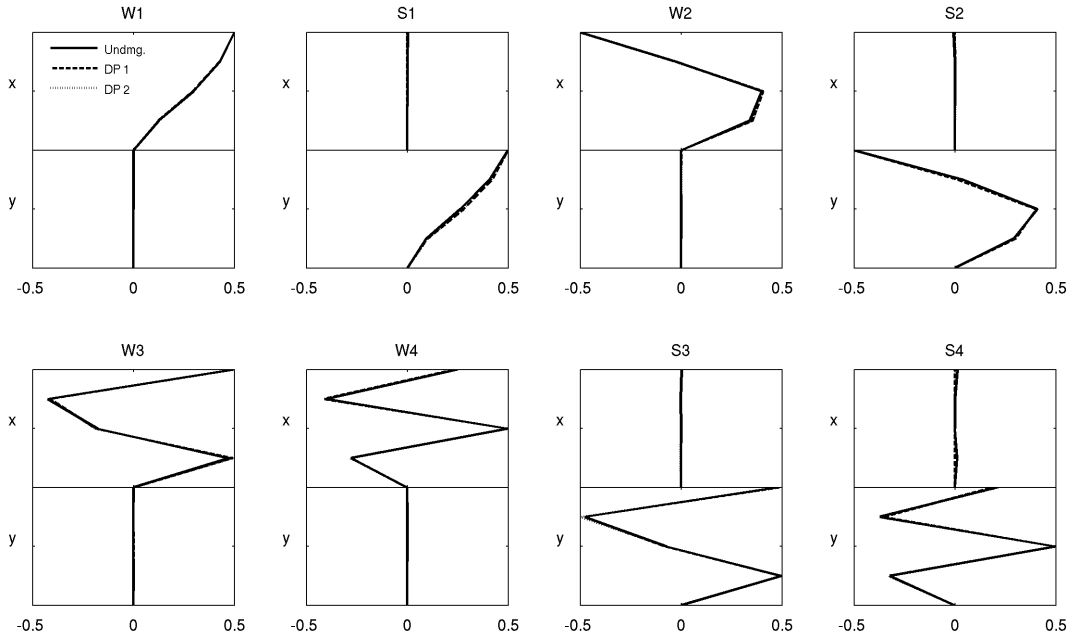


Figure 5.14: Sample means of the translational components of the system modeshapes for all damage cases.

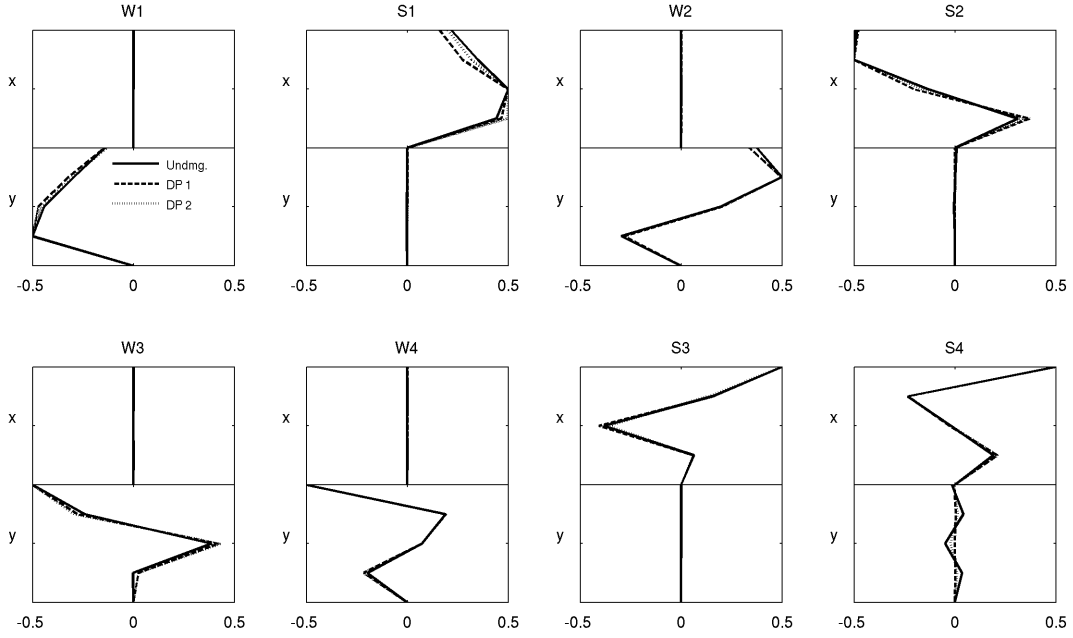


Figure 5.15: Sample means of the rotational components of the system modeshapes for all damage cases.

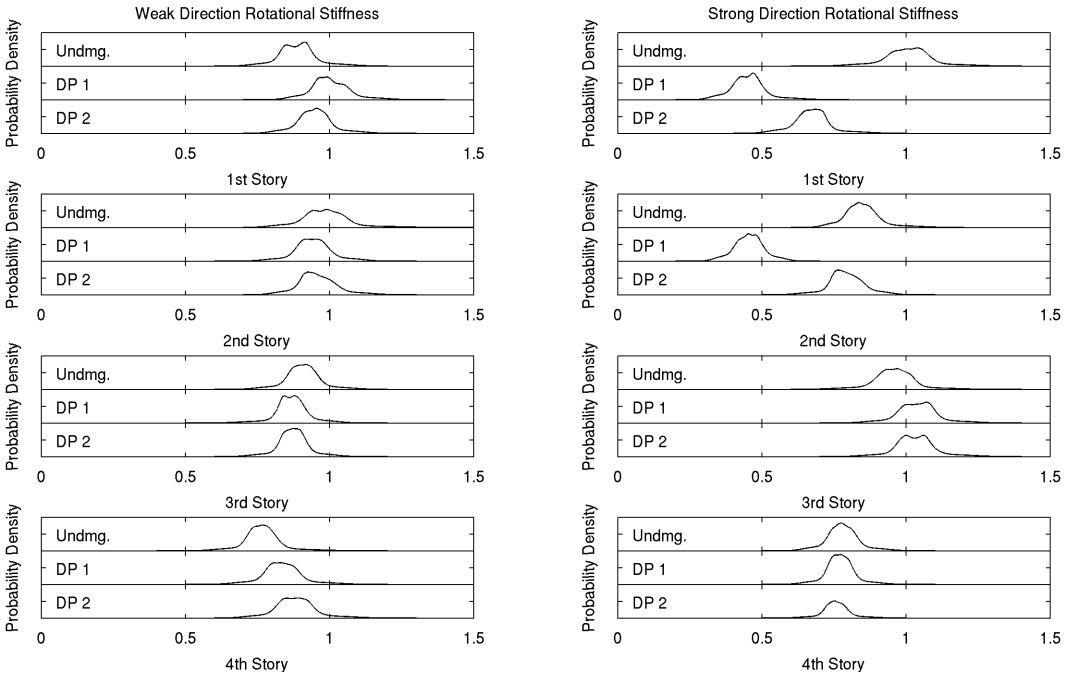


Figure 5.16: Kernel probability densities for the rotational stiffness parameters built from Markov chain samples.

probability densities (Silverman, 1986) for each parameter built from the combined post-burn-in samples. A region of support of 2000 samples was used. The damaged locations are clearly identified, and the stiffness parameters associated with undamaged structural elements generally show little difference between the PDFs generated using the undamaged and the damaged data. As for the 2-DOF case, damage probability curves were generated for the rotational stiffness parameters. Figures 5.17 and 5.18 shows these curves for DP 1 and DP 2, respectively. The damage curves are generated using samples for each damage configuration. For DP 1, convergence for the damage curve appears to require only 2000 post-burn-in samples, while for DP 2, between 3000 and 4000 samples are required.

To examine the variability of the damage probability curves if only one set of samples is generated for the undamaged and possibly damaged case, damage probability curves were generated using all 25 combinations of undamaged and damaged records for each damage pattern (since there are five runs available for each case). The mean damage probability curve and the mean damage probabilty curve plus and

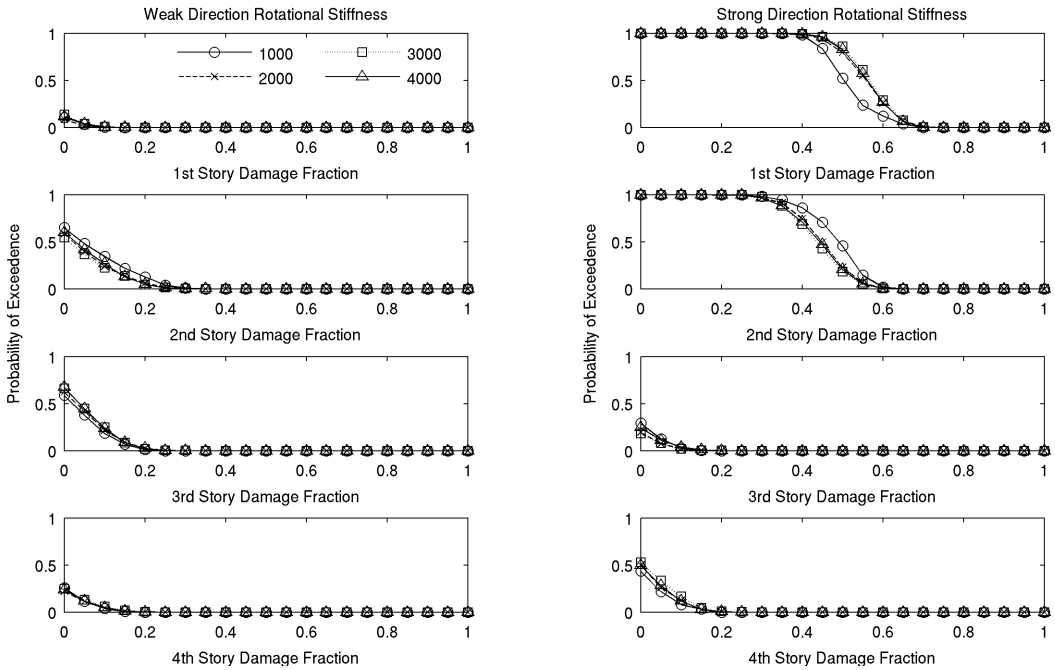


Figure 5.17: Damage probability curves for DP 1 using 1000, 2000, 3000, and 4000 post-burn-in samples

minus one standard deviation are shown in Figures 5.19 and 5.20. The median damage fraction for each stiffness parameter is calculated for each of the 25 damage probability curves, as in the previous section. The mean and standard deviation of the median damage fractions are listed in Table 5.7.

5.5 Conclusions

Thoughtful manipulation of the modal equations of a linear system leads to a form where the GS algorithm can be readily applied to update the structural model parameters, system modeshapes, and prediction-error variances. Comparison of the posterior PDFs of structural-model stiffness parameters, updated with data from undamaged and damaged systems, leads to a probabilistic description of the damage. This method has the advantage of providing the uncertainties associated with identified model parameters and with the damage estimate.

A simple 2-DOF shear-building example demonstrated the ability of the GS algo-

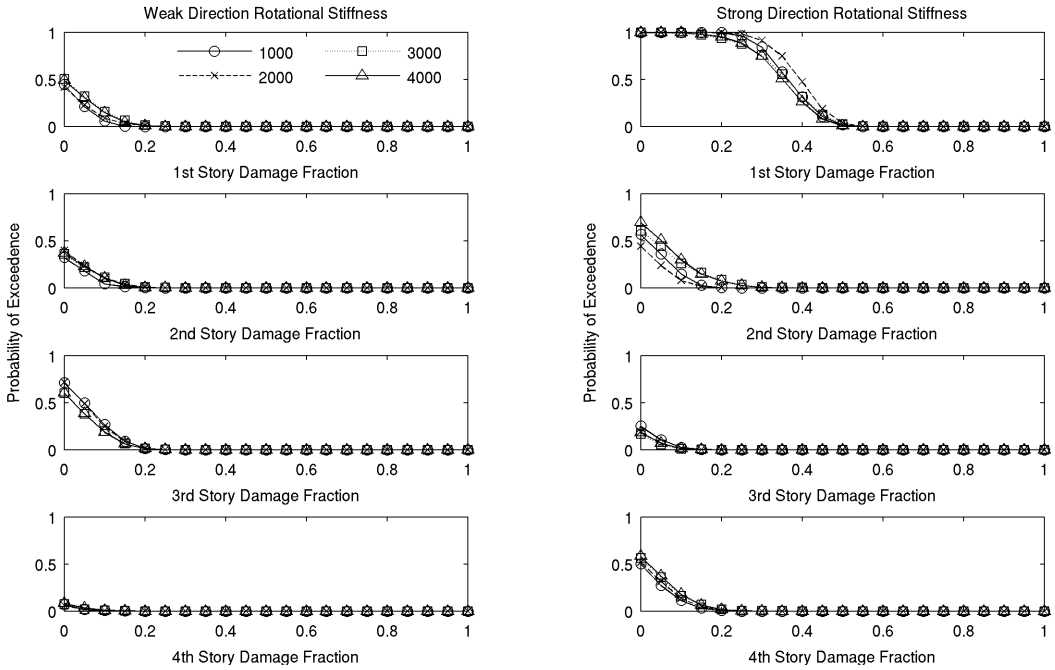


Figure 5.18: Damage probability curves for DP 2 using 1000, 2000, 3000, and 4000 post burn-in samples

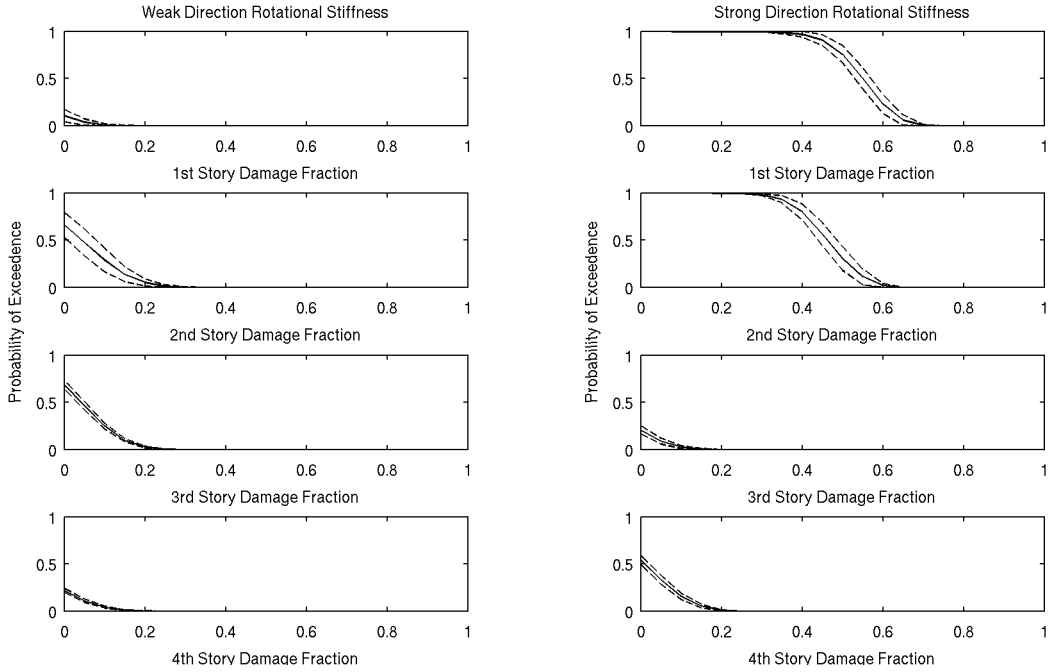


Figure 5.19: Mean damage probability curves for DP 1 using five Markov chains for the damaged and undamaged cases. The dashed lines represent the mean damage curve plus and minus one standard deviation.

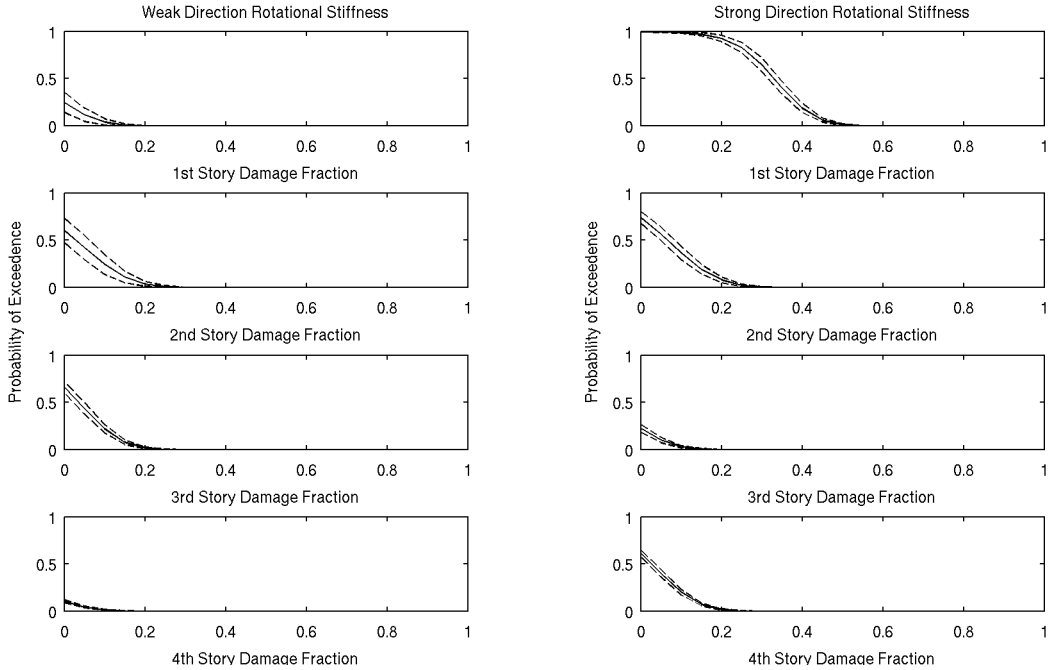


Figure 5.20: Mean damage probability curves for DP 1 using five Markov chains for the damaged and undamaged cases. The dashed lines represent the mean damage curve plus and minus one standard deviation.

Damage Pattern	Story	Mean Damage Fraction	
		Weak	Strong
DP 1	1^{st}	0.000 ± 0.000	0.550 ± 0.022
	2^{nd}	0.044 ± 0.033	0.464 ± 0.025
	3^{rd}	0.043 ± 0.010	0.000 ± 0.000
	4^{th}	0.000 ± 0.000	0.011 ± 0.010
DP 2	1^{st}	0.000 ± 0.000	0.331 ± 0.015
	2^{nd}	0.031 ± 0.029	0.067 ± 0.017
	3^{rd}	0.037 ± 0.013	0.000 ± 0.000
	4^{th}	0.000 ± 0.000	0.028 ± 0.008

Table 5.7: Median damage fractions for benchmark case. Values represent the mean and standard deviations.

rithm to deal with unidentifiable systems and also served to illustrate the effect that the prior PDFs can have on Bayesian model-class selection results.

The two-stage SHM methodology using the GS algorithm was successfully applied in two cases from the IASC-ASCE Structural Health Monitoring Simulated Benchmark problem. The cases studied are considered to be among the most challenging available for this benchmark, as the damage is associated with degrees of freedom that are not directly observed in the modeshapes identified from the data. The damage in each case was properly identified and localized. It is noted that the identification model included over 300 unknown parameters, demonstrating the ability of the GS algorithm to operate effectively in high-dimensional parameter spaces.

Future work may include applications of the methodology to a newly introduced benchmark problem for health monitoring of bridges (Caicedo et al., 2006), and to real data. The latter is expected to be challenging, due to observed shifts in the natural frequency due to environmental changes that may prove difficult to distinguish from structural damage (Clinton et al., 2006).

Chapter 6

Hysteretic Structural Models

Identification of structural models from earthquake response can play a key role in structural health monitoring, structural control, and improving performance-based design. However, implementation is complicated by the non-linear response of structures under strong seismic loading; in particular, the structural restoring forces are hysteretic, depending on the previous time history of the structural response rather than an instantaneous finite-dimensional state.

Current methods for developing finite-element models can produce structural responses that are qualitatively consistent with behavior observed during strong earthquake shaking, but system-identification methods based on updating these finite-element models with measured seismic responses are challenging, because the large number of uncertain parameters associated with these models makes the inverse problem extremely ill conditioned.

Simplified models can be used in the identification procedure but the selection of an appropriate class of models to employ is complicated by the hysteretic structural restoring forces. Although some research into the identification of hysteretic systems has been carried out (Jayakumar and Beck, 1988; Cifuentes and Iwan, 1989; Benedettini et al., 1995), this earlier work does not quantify the modelling uncertainties. A Bayesian updating approach has the advantage of being able to quantify the uncertainty associated with model prediction of system response and to handle ill-conditioned problems. In particular, stochastic simulation methods seem to be particularly well suited to the problem of model updating and model-class selection

for such systems.

6.1 Masing Hysteretic Models

One fundamental approach to constructing hysteretic force-deformation relations for structural members or assemblages of structural members is to build them up from constitutive equations (plasticity models), which govern material behavior at a point. However, factors such as complex stress distributions, material inhomogeneities, and the large number of structural elements make this approach impractical. Also, there is no general consensus on the choice of plasticity model under arbitrary loading.

An alternative approach is to develop simplified models that capture the essential features of the hysteretic force-deformation relationship and then to validate the models against the observed behavior of structures. This has been done by Jayakumar (1987) for the well-known Bouc-Wen model (Wen, 1976), which is essentially a planar version of the early endochronic model (Valanis 1971); these models are mathematically convenient, especially for random-vibration studies using equivalent linearization; but when they are subjected to asymmetric cyclic loading, these models can exhibit an unphysical "drifting" behavior (Jayakumar, 1987), as shown in Figure 6.1. This behavior makes this class of models unsuitable for identification using strong seismic responses, where irregular loading occurs.

A simplified hysteretic model with a physical basis was presented by Masing (1926), which is based on the hypothesis that a one-dimensional hysteretic system

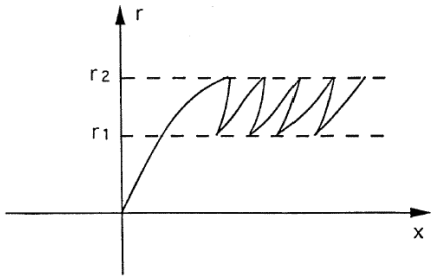


Figure 6.1: Drifting behavior for Bouc-Wen model under asymmetric cyclic loading.

could be viewed as a collection of many ideal elasto-plastic elements (a linear spring in series with a Coulomb damper) with the same elastic stiffness but different yield strengths. This idea was used in structural dynamics to form the distributed dlement model (DEM), which consists of a finite or infinite collection of N ideal elasto-plastic elements connected in parallel (Iwan, 1966, 1967) with a common stiffness k/N for the springs but with different yield strengths $r_i^*/N, i = 1, \ldots, N$, as shown in Figure 6.2. The restoring force r for a single-degree-of-freedom DEM subjected to a displacement x is given by

$$r = \sum_{i=1}^{n} \frac{r_i^*}{N} + kx\frac{N-n}{N} \tag{6.1}$$

where n is the number of elements that have yielded. Infinite collections of elasto-plastic elements can be considered by introducing a yield-strength distribution function $\phi(r^*)$, such that the restoring force $r(x)$ during initial loading is

$$r(x) = \int_0^{kx} r^*\phi(r^*)dr^* + kx\int_{kx}^{\infty} \phi(r^*)dr^* \tag{6.2}$$

where the first and second terms represent the contribution to the restoring force of yielded and unyielded elements, respectively. Because there is an underlying physical basis for the model, DEMs with a finite number of elements have been shown to give good representations of the hysteretic behavior of some structures and do not exhibit the previously discussed drifting behaviors. However, DEMs with an infinite number of elements are difficult to implement directly, in contrast to the finite case where the state of each element is tracked, though efficient algorithms for these types of DEMs are the focus of recent work (Ashrafi et al., 2005). Fortunately, there are two hysteretic rules that exactly describe the behavior of DEMs without needing to track the internal behavior of the elements, which are now presented.

Masing (1926) postulated that the steady-state behavior of such a hysteretic system subjected to cylic loading could be described as follows. If the initial or "virgin" loading curve is described by the implicit relationship

$$f(x, r) = 0 \tag{6.3}$$

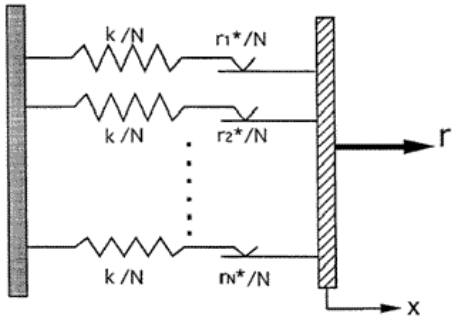


Figure 6.2: Conceptual sketch of the distributed element model. Taken from Chiang (1992).

where x is the deflection and r is the restoring force, then each branch of the hysteresis loop between the points (x_a, r_a) and $(-x_a, -r_a)$ is given by

$$f\left(\frac{x - x^*}{2}, \frac{r - r^*}{2}\right) = 0 \tag{6.4}$$

where (x^*, r^*) is the load-reversal point for that branch curve. For the unloading branch, $(x^*, r^*) \equiv (x_a, r_a)$ and $(x^*, r^*) \equiv (-x_a, -r_a)$ for the loading branch. This is commonly referred to as Masing's rule.

Masing's theory was extended to apply to the case of softening hysteretic systems under arbitrary loading by Jayakumar (1987), who specified two hysteresis rules, which will henceforth be referred to as the extended Masing rules:

1. The equation of any hysteretic force-deformation curve can be obtained by applying the original Masing rule to the virgin loading curve using the latest point of load reversal. For example, if the virgin loading curve OA in Figure 6.3 is given by Equation 6.3, then the branch curve CD is defined by Equation 6.4, with $(x^*, r^*) \equiv (x_c, r_c)$.

2. Once an interior curve under continued loading or unloading crosses a curve from a previous load cycle, a hysteresis loop is completed and the load-deformation curve then follows that of the previous cycle. For example, if the unloading curve DE in Figure 6.3 is continued to point C, further unloading will follow a

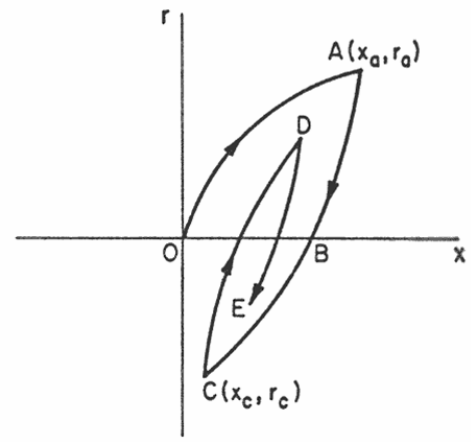


Figure 6.3: Hysteretis loop for transient loading of extended Masing model (Beck and Jyakumar 1996).

path that is the extension of the curve ABC.

As already mentioned, the class of models defined by these rules was shown by Jayakumar (1987) to describe the behavior of DEMs, which have an underlying physical model, and so they are expected to produce reasonable hysteretic behavior. Chiang (1992) later demonstrated the relationship between the initial loading curve $f(x, r)$ in Equation 6.3 and the yield-strength distribution function $\phi(r^*)$ in Equation 6.2 as follows:

$$\frac{dr}{dx} = k \int_{kx}^{\infty} \phi(r^*) dr^* \tag{6.5}$$

which also implies that the yield-strength distribution function may be calculated from the initial loading curve:

$$\phi(kx) = -\frac{1}{k^2} \frac{d^2 r}{dx^2} \tag{6.6}$$

6.1.1 Masing Shear-Building Model

Jayakumar and Beck (1988) performed system identification of a full-scale six-story steel building tested pseudo-dynamically in the laboratory of the Building Research Institute in Tsukuba, Japan (Foutch et al., 1986). Substantial yielding occurred in the lower stories during the test, which was an experimental simulation under the Taft

record from the 1952 Kern County earthquake in California. For the identification, Jayakumar and Beck used a class of shear-building models in which a Masing model was used to describe the relationship between the story shear forces and the inter-story drifts. However, multiple optimal models were obtained when using a least-squares output-error approach, with the results depending heavily on the initial choice of the model parameters. This suggests that updating this class of models is an ill-conditioned problem well suited for study with stochastic simulation technqiues.

Consider a structural model with a rigid foundation where the vector of relative displacements of the structure $x(t)$ is related to the ground acceleration $\ddot{y}(t)$ as follows:

$$M\ddot{x} + C\dot{x} + R = -Mb\ddot{y} \tag{6.7}$$

where M is the mass matrix, C is the viscous-damping matrix, R is the vector of restoring forces acting on each floor mass and b is the pseudo-static influence vector. The restoring force at the ith floor mass is given by:

$$R_i = r_i - r_{i+1} \tag{6.8}$$

where r_i the inter-story shear force at the ith story, which for initial loading is related to the inter-story drifts by

$$\dot{r}_i = K_i(\dot{x}_i - \dot{x}_{i-1}) \left[1 - \left|\frac{r_i}{r_{u,i}}\right|^{\alpha_i}\right] \tag{6.9}$$

where K_i is the initial story-stiffness term, $r_{u,i}$ is the story ultimate strength, and the smoothness of the transition from elastic to plastic response is controlled by the positive parameter α_i, as shown in Figure 6.4. Note that for $i = n$ in Equation 6.8, $r_{n+1} = 0$; and for $i = 1$ in Equation 6.9, $x_0 = 0$. This class of models represents a sub-class of Masing models; through the choice of initial loading curves, a huge variety of hysteretic models can be described by the two extended Masing rules, including many existing hysteretic models (Jayakumar, 1987).

The basic form of the force-deformation relationship given in Equation 6.9 is

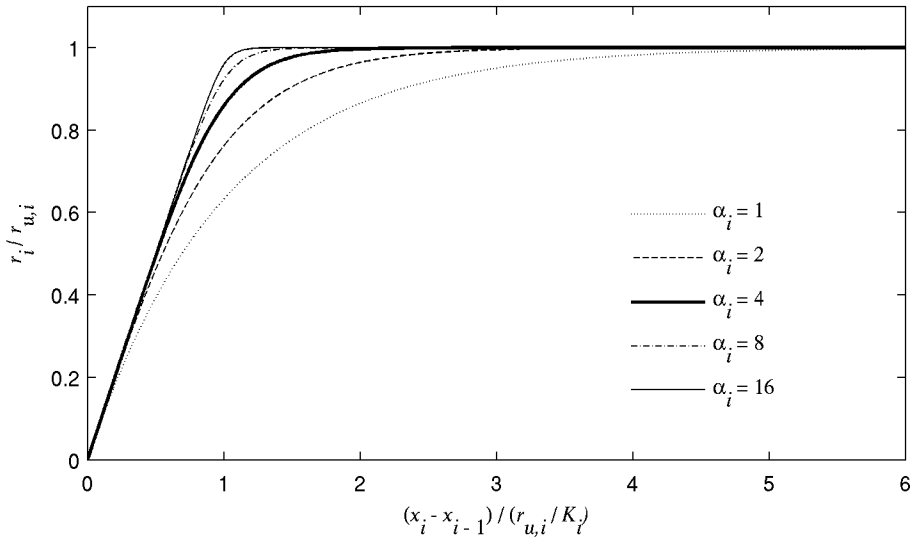


Figure 6.4: Plots of initial loading curves for Masing shear-building model for different values of the elastic-to-plastic transition parameter α_i.

similar to that of a Bouc-Wen model. The major difference is that this equation serves only as the initial loading curve and must be combined with the two extended Masing rules, rather than giving a complete description of the structural response.

6.1.2 Specifying Hysteretic Models with the Yield-Strength Distribution Function

The implied yield-strength distribution functions for the Masing shear-building model can be derived from Equation 6.6. The distribution functions for different values of elastic-to-plastic transition parameters are shown in Figure 6.5. The resulting distributions appear to be plausible distributions for element yield strengths, with the spread of the strengths decreasing with the "sharpness" of the transition between elastic and plastic behavior. Of particular note are the special cases $\alpha_i = 1$, which results in an exponential distribution, and $\alpha_i = 2$, which produces a lognormal distribution. Consequently, the force-displacement relationship for these two cases can be determined explicitly (Jayakumar, 1987).

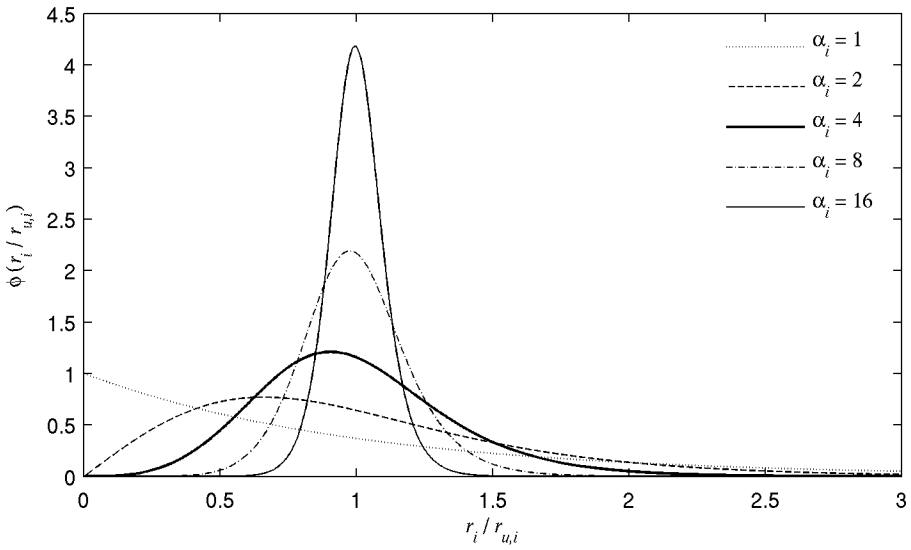


Figure 6.5: Plots of yield-strength distribution functions for Masing shear-building model for different values of the elastic-to-plastic transition parameter α_i.

Chiang (1992) studied a class of models defined by what he called a generalized Rayleigh distribution on the yield strength. Specifically:

$$\phi(r) = \frac{\eta}{r}\left(\Gamma\left(\frac{\eta+1}{\eta}\right)\frac{r}{r_u}\right)^{\eta}\exp\left[-\left(\Gamma\left(1+\frac{1}{\eta}\right)\frac{r}{r_u}\right)^{\eta}\right] \tag{6.10}$$

Generalized Rayleigh distributions are plotted for different values of the positive elastic-to-plastic transition parameter η in Figure 6.6. The differential formulation of the force-displacement relationship can be expressed as

$$\frac{dr}{dx} = K\exp\left[-\left(\Gamma\left(\frac{\eta+1}{\eta}\right)\frac{Kx}{r_u}\right)^{\eta}\right] \tag{6.11}$$

where the gamma function $\Gamma(.)$ is defined as

$$\Gamma(a) = \int_0^{\infty} e^{-t}t^{a-1}dt \tag{6.12}$$

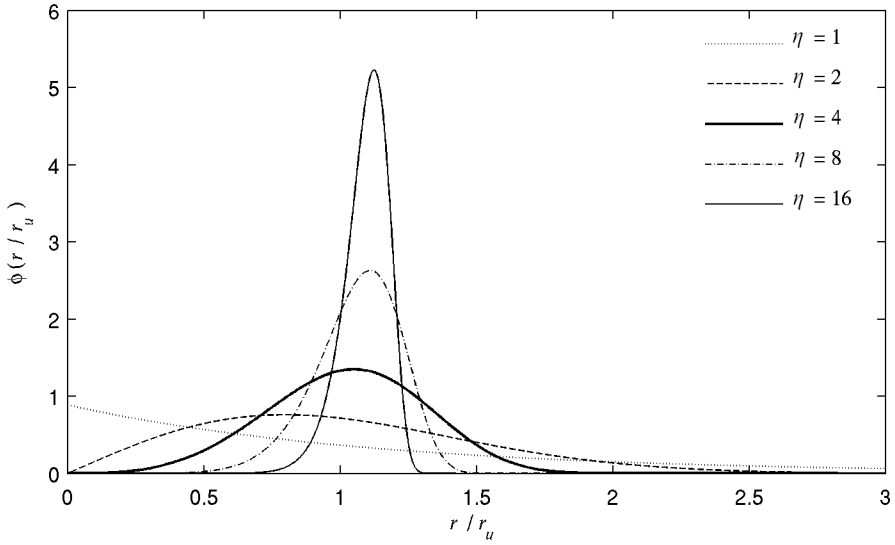


Figure 6.6: Plots of generalized Rayleigh distribution for yield strength for different values of the elastic-to-plastic transition parameter η.

The advantage of this distribution is that the explicit form of the force-displacement relationship can be calculated:

$$r(x) = r_u \Gamma\left(\nu x^\eta, \frac{\eta+1}{\eta}\right) + Kx\exp\left(-\nu x^\eta\right) \tag{6.13}$$

where $\nu \equiv \left(\Gamma\left(\frac{\eta+1}{\eta}\right)\frac{K}{r_u}\right)^\eta$ and $\Gamma(.,.)$ is the incomplete gamma function:

$$\Gamma(x,a) = \frac{1}{\Gamma(a)}\int_0^x e^{-t}t^{a-1}dt \tag{6.14}$$

The initial loading curves for this model, which will be referred to as generalized Rayleigh models, for different values of the elastic-to-plastic transition parameters η are shown in Figure 6.7.

6.1.3 Developing Reasonable Prior PDFs for Masing Models

Recall from Chapter 2 that, in general, for globally identifiable systems with large amounts of data, the prior PDFs do not significantly affect the results of model parameter updating (though they can play a critical role in model-class selection).

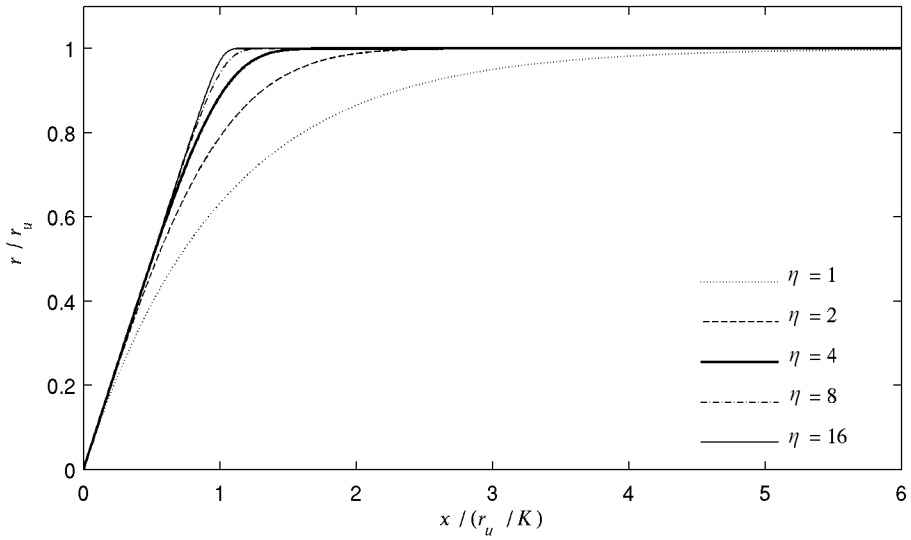


Figure 6.7: Plots of initial loading curves for generalized Rayleigh model for different values of the elastic-to-plastic transition parameter η.

However, for ill-conditioned problems, the prior PDFs can have a great influence on the final estimate, making the selection of reasonable prior PDFs an important consideration.

In this respect, the Masing shear-building model presented previously has an important advantage; most of the model parameters correspond to actual properties (the initial stiffness and ultimate strength) and initial estimates can be calculated from material properties and structural drawings (Jayakumar, 1987).

Prior probability distributions for the mass parameters may be taken as relatively peaked if structural drawings of the building design are available. If not, any information about the type of construction materials and the use of the building can be used to make an informed estimate (with correspondingly higher uncertainties).

If information on the member properties is available, the mean value of the linear story stiffnesses can be calculated from finite elements or using an empirical method such as Biggs' formula (Anagnostopoulos et al., 1972). However, it is also possible to rapidly obtain an estimate of the linear stiffnesses of a shear-building system using only the modal properties.

The first-mode approximation method (Nielsen, 1964; Lai and Vanmarcke, 1980)

can be used to estimate the story stiffnesses as follows: If the first mode has mode-shape components $\phi_1^T = [\phi_{1,1}\phi_{1,2}\ldots\phi_{1,N_d}]$ then

$$K\phi_1 = \Phi \begin{bmatrix} k_1 & k_2 & \cdots & k_{N_d} \end{bmatrix}^T \tag{6.15}$$

$$\Phi = \begin{bmatrix} \phi_{1,1} & \phi_{1,1}-\phi_{1,2} & 0 & 0 & \cdots & 0 \\ 0 & -\phi_{1,1}+\phi_{1,2} & \phi_{1,2}-\phi_{1,3} & 0 & \cdots & 0 \\ 0 & 0 & -\phi_{1,2}+\phi_{1,3} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 \\ \vdots & \vdots & & & \ddots & \phi_{1,Nd-1}-\phi_{1,Nd} \\ 0 & 0 & \cdots & \cdots & 0 & -\phi_{1,Nd-1}+\phi_{1,Nd} \end{bmatrix} \tag{6.16}$$

Recall from Equation 5.4 that $K\phi_1 = \omega_1^2 M\phi_1$, so given an estimate of the first mode shape and natural frequency, as well as the mass matrix, the story stiffnesses can be approximated. Note that the matrix defined in Equation 6.16 is singular (non-invertible) if the first component of the modeshape, $\phi_{1,1}$, is zero, or if any two consecutive modeshape components are equal, that is, $\phi_{1,i} = \phi_{1,i+1}$, $i = 1,\ldots,N_d - 1$. As the relative displacement approaches zero, the associated stiffness approaches infinity.

6.2 Studies with Simulated Data

To investigate the performance of the TMCMC algorithm with the class of Masing models for identification of hysteretic systems, an initial study was performed using simulated dynamic response data. The data is generated using a three-story Masing shear-building system described by Equations 6.3 through 6.9 and the two extended Masing rules. Each story has a mass of 1.25×10^5 kg and a small-amplitude inter-story stiffness of $K_i = 2.5\times10^8$ N/m. The ultimate strength of each story is $r_{u,i} = 1.75\times10^6$ N, and the value of the elastic-to-plastic transition parameter for each story is $\alpha_i = 4$.

The natural frequencies of the structure (based on a linear model given by the small-amplitude stiffnesses) are 3.17 Hz, 8.88 Hz, and 12.86 Hz. The structure is

excited with the Sylmar ground motion recorded at Olive View Hospital during the 1994 Northridge earthquake. The resulting response has peak drifts of 2.77 cm, 0.85 cm, and 0.50 cm for the first, second, and third stories, respectively. Viscous damping is provided to the system through a Rayleigh damping matrix (sum of a term proportional to the mass matrix M and a term proportional to the small-amplitude stiffness matrix K) with coefficients $c_M = 0.293$ and $c_K = 2.64 \times 10^{-4}$ chosen to provide 1% of critical damping in the first and second modes.

Two sets of responses are simulated to provide the data for model updating. The first set corresponds to the time histories of inter-story drift, referred to henceforth as $\mathcal{D}_{drift}$. To simulate measurement noise, a small amount of Gaussian discrete white noise is added to each channel of generated data, with a standard deviation of 0.1 cm for each measurement. This simulated data is shown in Figure 6.8. Each channel corresponds to 500 data points at a sampling interval of 0.02 seconds. While dynamic measurements of inter-story drifts are possible and have been performed for laboratory testing of structures, their measurement in actual structures is rare. Accelerometers are more commonly used for this purpose, so the second set of data, $\mathcal{D}_{acc}$, consists of simulated absolute accelerations of each floor. Again, Gaussian discrete white noise is added to the responses, this time with a standard deviation of 0.5 m/s^2, which is equal to about 20% of the root-mean-square value of the time histories. Simulated accelerations are plotted in Figure 6.9. The hysteresis loops generated by the Sylmar ground motion for each story are shown in Figure 6.10. As indicated by the figure, yielding in the simulated structure occurs primarily in the first story, with some moderately non-linear behavior in the second story and almost linear response for the third story.

6.2.1 Identification Model Classes

Four model classes are considered for system identification. All of them use the Masing shear-building model in Equations 6.7 through 6.9 to generate the predicted response, $q_t^{(i)}, i = 1, \ldots, N_d, t = 1, \ldots, N_t$, for $N_d = 3$ channels of $N_t = 500$ time-

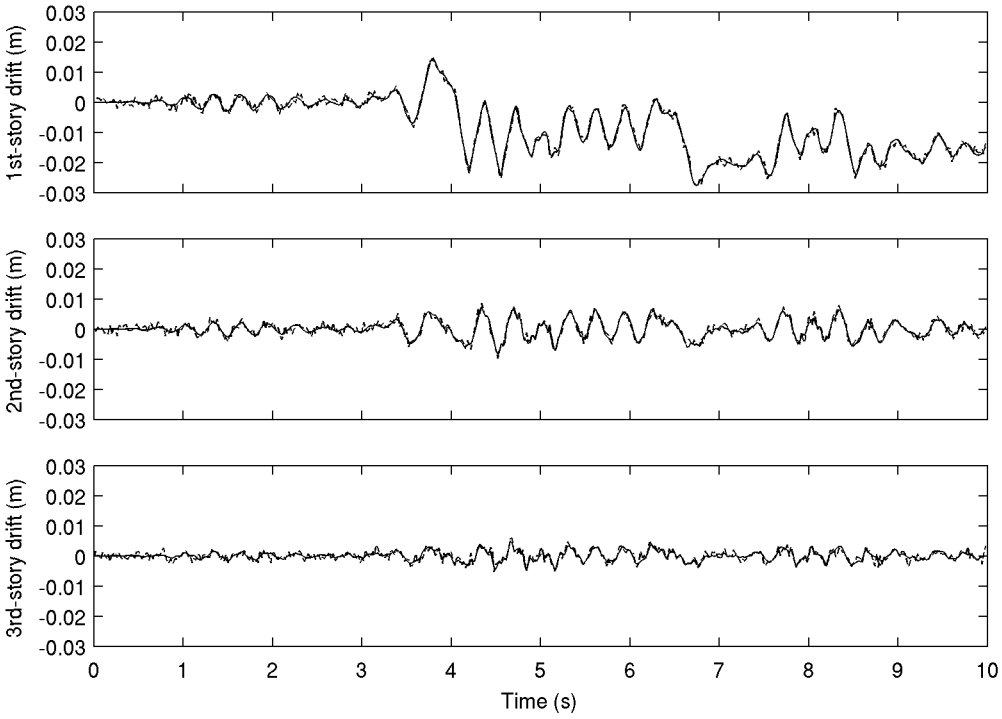


Figure 6.8: Simulated inter-story drift time histories.

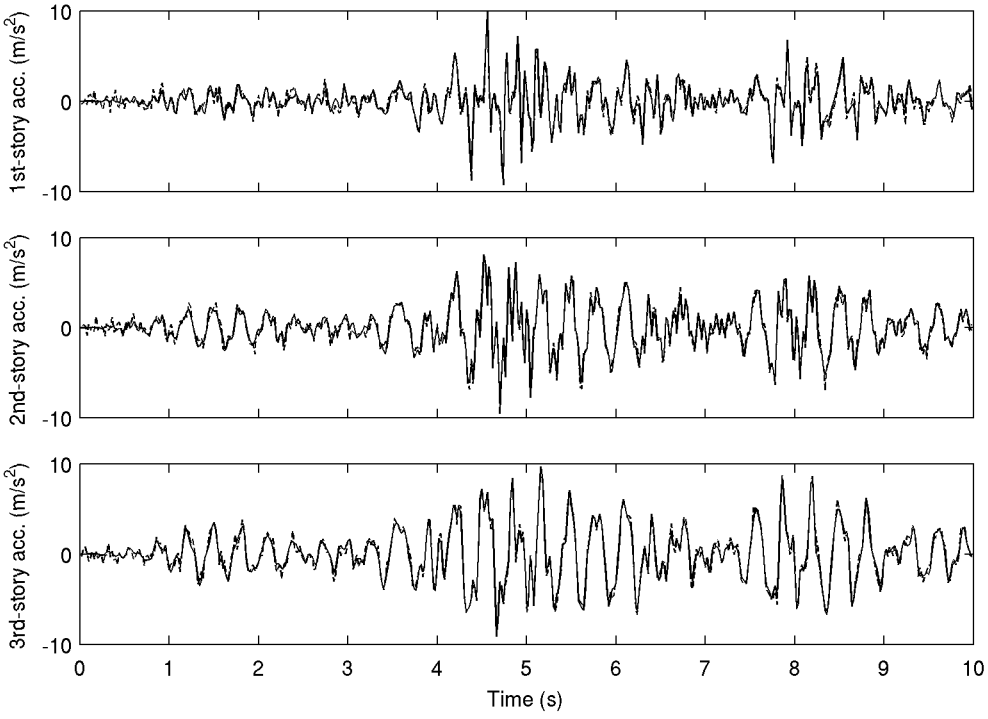


Figure 6.9: Simulated floor acceleration time histories.

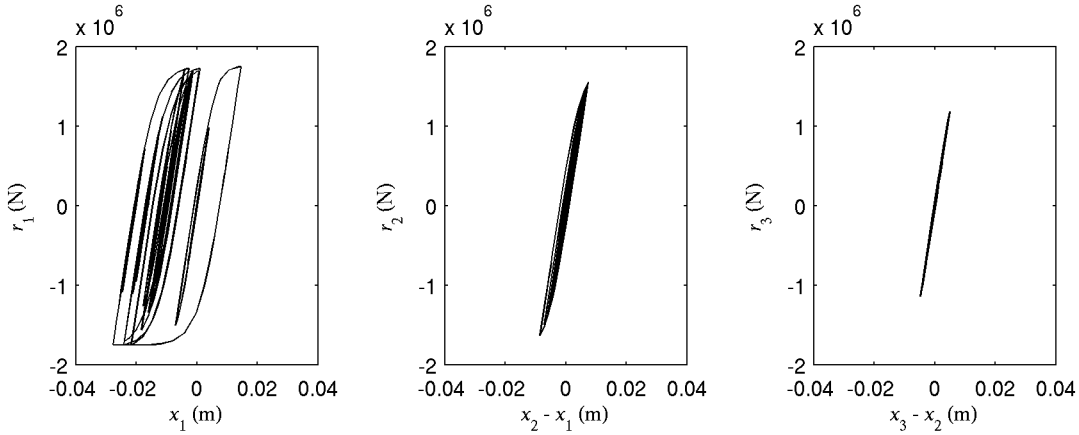


Figure 6.10: Simulated inter-story shear forces plotted against inter-story drifts.

points at a time-step of 0.02 seconds. The vector θ of uncertain model parameters include those defining the viscous-damping matrix C; that is, c_K and c_M, and the nine hysteretic structural parameters K_i, $r_{u,i}$, and α_i, $i = 1, 2, 3$, from Equation 6.9. The prediction error is taken as Gaussian with zero mean and an equal (uncertain) variance σ_q^2 for each channel. Therefore, the likelihood function is

$$p(\mathcal{D} \mid \theta, \mathcal{M}) = \frac{1}{(2\pi\sigma_q^2)^{\frac{N_t N_d}{2}}} \exp\left[-\frac{1}{2\sigma_q^2} \sum_{i=1}^{N_d} \sum_{t=1}^{N_t} \left(q_t^{(i)}(\theta) - \hat{q}_t{}^{(i)}\right)^2\right] \tag{6.17}$$

where $\hat{q}_t{}^{(i)}$ is the measurement for channel i at time-point t.

The mass matrix M is equal to the actual mass matrix, which is reasonable since, as discussed, the mass distribution can be accurately evaluated based on structural drawings. The three small-amplitude stiffnesses and ultimate strengths of the three stories are to be estimated, along with the elastic-to-plastic transition parameters α_i, which are constrained to be equal for all three stories in model classes $\mathcal{M}_1$ and $\mathcal{M}_2$ but are allowed to vary between stories for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$. Rayleigh damping coefficients are estimated for model classes $\mathcal{M}_2$ and $\mathcal{M}_4$, as described below.

The prior probablity distributions for the small-amplitude stiffnesses K_i are taken to be independent lognormal distributions, each with logarithmic mean $\ln(2.5 \times 10^8)$ and logarithmic standard deviation of 0.5. The PDFs for the ultimate strengths $r_{u,i}$ are also independent lognormal distributions with logarithmic means of $\ln(2.5 \times 10^6)$

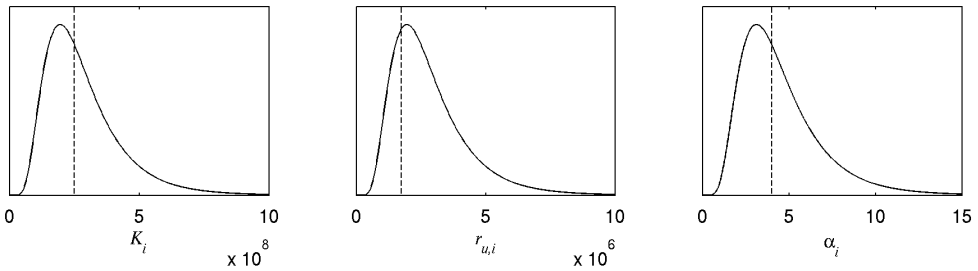


Figure 6.11: Prior PDFs for Masing shear-building model parameters. The dashed lines indicate the values of the parameters used to generate the data.

and a logarithmic standard deviations of 0.5. The elastic-to-plastic transition parameter α_i is also lognormally distributed with logarithmic mean of $\ln(4)$ and a logarithmic standard deviation of 0.5. Plots of the prior PDFs for K_i, $r_{u,i}$, and α_i are shown in Figure 6.11.

The prediction-error variances for the identification with drift records is assumed to be equal for all three stories and the prior PDF for σ^2_{drift} is uniformly distributed between 0 and 2.5×10^{-5}, which is approximately one-half of the mean square of the "measured" drift time histories. Similarly, the prediction-error variance for identification with acceleration records, σ^2_{acc}, is equal for all the stories and is uniformly distributed between 0 and 3, which is again approximately one-half the mean square of the "measured" acceleration time histories.

Model classes $\mathcal{M}_1$ and $\mathcal{M}_3$ contain no Rayleigh viscous-damping matrix, whereas model classes $\mathcal{M}_2$ and $\mathcal{M}_4$ do, and so damping coefficients c_K and c_M are included as uncertain parameters to be updated for those two model classes. The prior PDF for c_M is a uniform distribution between 0 and 1.5, and the prior PDF for c_K is a uniform distribution between 0 and 1.5×10^{-3}. For both coefficients, the range of the prior PDF is more than five times the actual value used to generate the data. It should be noted that the model used to generate the data is contained in model classes $\mathcal{M}_2$ and $\mathcal{M}_4$. The candidate model classes are summarized in Table 6.1.

Model Class	K_1	K_2	K_3	$r_{u,1}$	$r_{u,2}$	$r_{u,3}$	α_1	α_2	α_3	c_M	c_K	σ
$\mathcal{M}_1$	free	free	free	free	free	free	free	$=\alpha_1$	$=\alpha_1$	fixed	fixed	free
$\mathcal{M}_2$	free	free	free	free	free	free	free	$=\alpha_1$	$=\alpha_1$	free	free	free
$\mathcal{M}_3$	free	free	free	free	free	free	free	free	free	fixed	fixed	free
$\mathcal{M}_4$	free	free	free	free	free	free	free	free	free	free	free	free

Table 6.1: Summary of parameters for candidate model classes.

6.2.2 Model Updating and Model-Class Selection with Drift Time Histories

For each model class, three runs were performed using the TMCMC algorithm. Each run used 1000 samples per level. Between 19 to 23 levels were needed to go from the prior PDF to the posterior PDF in a given run. Figure 6.12 shows how the samples in the $(r_{u,1}, K_1)$ space converge for a run using model class $\mathcal{M}_1$ as the tempering parameter β increases, where the model class is globally identifiable. Figure 6.13 shows the convergence of samples in the $(r_{u,3}, K_3)$ space for model class $\mathcal{M}_2$, where the model class is close to unidentifiable in the $r_{u,3}$ direction, because of the lack of third-story yielding in the data, as discussed later. The estimated most probable values and standard deviations of the marginal posterior PDFs for each parameter, calculated from the aggregate samples of the three runs performed for each model class, are presented in Table 6.2. For purposes of comparison, the most probable parameter values for each model class are also calculated by numerical optimization of the posterior PDF, and also are shown in Table 6.2. The most probable values obtained by optimization are within one standard deviation of the corresponding values obtained by stochastic simulation, showing good agreement despite the expected discrepancy between maximizing the joint PDF and maximizing the marginal PDFs (discussed in Section 3.2). It should be noted that convergence of the optimization algorithm was achieved using initial conditions based on the simulation results, otherwise there would have been difficulties in the optimization due to ill-conditioning, particularly in model classes $\mathcal{M}_3$ and $\mathcal{M}_4$, as will be seen later. For model classes $\mathcal{M}_1$ and $\mathcal{M}_2$, the estimates of the three stiffness parameters are fairly well-constrained (e.g., see Figures 6.12 and 6.13). The estimate for the first-story strength is also very tightly

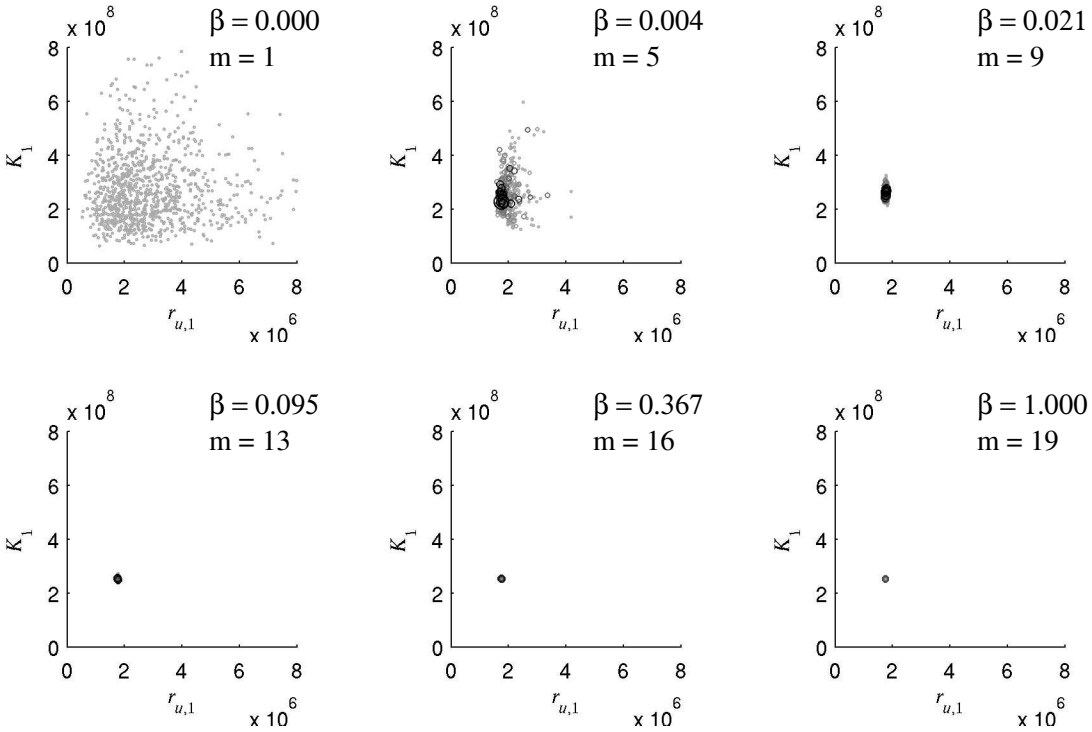


Figure 6.12: Plots of the samples in the $(r_{u,1}, K_1)$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_1$ with drift data. Repeated samples are indicated by size and shading of the markers.

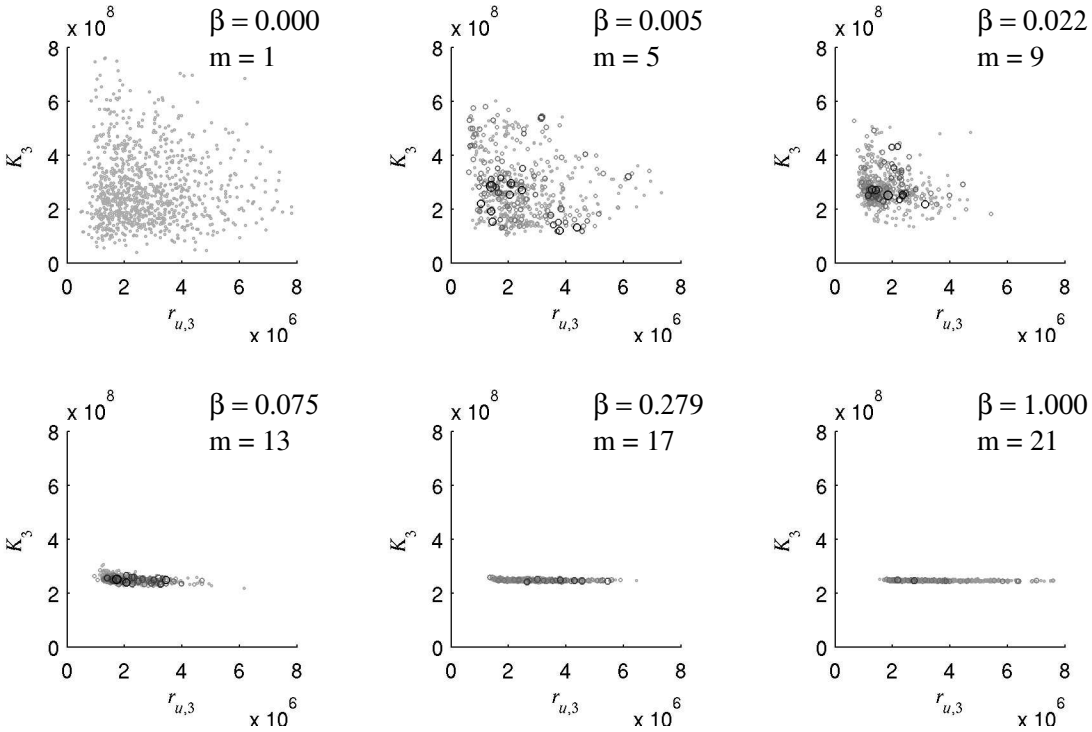


Figure 6.13: Plots of the samples in the $(r_{u,3}, K_3)$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_2$ with drift data. Repeated samples are indicated by size and shading of the markers.

Model Class	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$		$\mathcal{M}_4$	
	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.
K_1	2.523	2.525	2.501	2.504	2.531	2.525	2.510	2.507
10^8 N/m	(0.007)	[0.002]	(0.015)	[0.003]	(0.009)	[-0.006]	(0.017)	[-0.003]
K_2	2.503	2.500	2.486	2.488	2.477	2.483	2.468	2.472
10^8 N/m	(0.012)	[-0.003]	(0.021)	[0.002]	(0.019)	[0.006]	(0.026)	[0.004]
K_3	2.492	2.495	2.474	2.476	2.530	2.522	2.472	2.492
10^8 N/m	(0.013)	[0.003]	(0.017)	[-0.002]	(0.022)	[-0.008]	(0.018)	[0.020]
$r_{u,1}$	1.752	1.752	1.752	1.753	1.747	1.747	1.749	1.751
10^6 N	(0.005)	[0.000]	(0.004)	[0.001]	(0.005)	[0.000]	(0.005)	[0.002]
$r_{u,2}$	1.756	1.764	1.809	1.815	1.691	1.685	1.757	1.748
10^6 N	(0.036)	[0.008]	(0.039)	[0.006]	(0.066)	[-0.006]	(0.066)	[-0.009]
$r_{u,3}$	1.736	1.731	2.480	2.061	2.725	2.634	2.714	2.582
10^6 N	(0.129)	[-0.005]	(0.901)	[-0.419]	(0.972)	[-0.091]	(1.140)	[-0.132]
α_1	3.916	3.890	3.911	3.922	3.899	3.949	3.879	3.891
	(0.110)	[-0.026]	(0.103)	[0.011]	(0.110)	[0.060]	(0.098)	[0.012]
α_2	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	4.513	4.542	4.350	4.479
	(fixed)	(fixed)	(fixed)	(fixed)	(0.643)	[0.029]	(0.496)	[0.129]
α_3	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	4.513	4.542	4.350	4.479
	(fixed)	(fixed)	(fixed)	(fixed)	(0.643)	[0.029]	(0.496)	[0.129]
c_M	0.000	0.000	0.293	0.294	0.000	0.000	0.261	0.299
	(fixed)	(fixed)	(0.049)	[0.001]	(fixed)	(fixed)	(0.061)	[0.038]
c_K	0.000	0.000	3.269	3.021	0.000	0.000	3.474	3.073
10^{-4}	(fixed)	(fixed)	(0.660)	[-0.248]	(fixed)	(fixed)	(1.156)	[-0.401]
σ^2_{drift}	1.139	1.129	1.021	1.009	1.130	1.125	1.032	1.008
10^{-6} m^2	(0.044)	[-0.010]	(0.041)	[-0.012]	(0.043)	[-0.005]	(0.036)	[-0.024]

Table 6.2: The estimated most probable parameter values and standard deviations for the marginal PDFs, calculated from samples, and for the joint PDF, calculated by optimization of the posterior, for model updating with drift data using both stochastic simulation and numerical optimization. Standard deviations are shown in parentheses, and the difference between the parameter values maximizing the marginal and joint PDFs is shown in brackets.

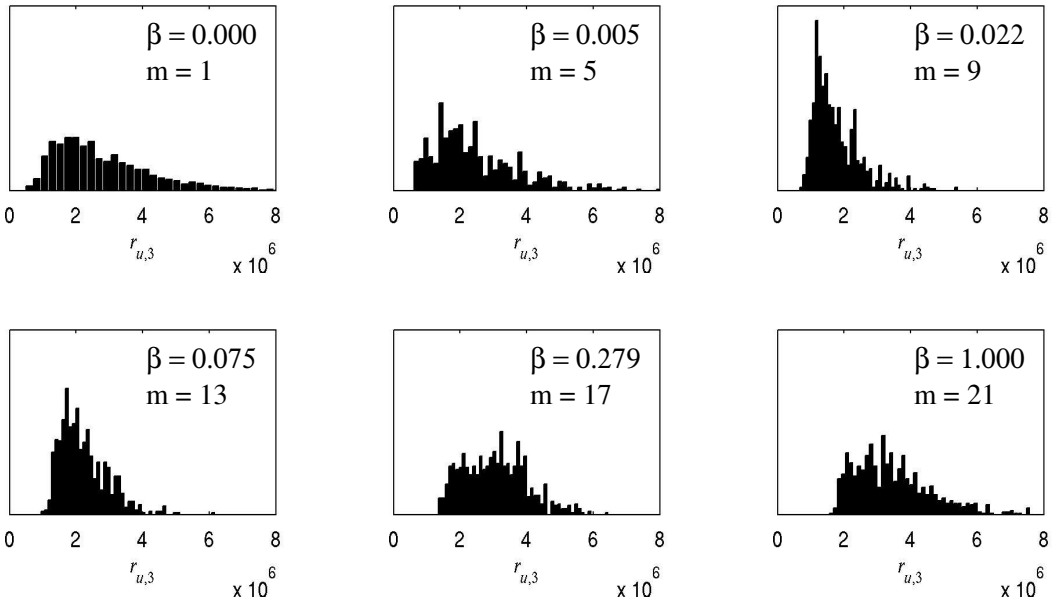


Figure 6.14: Normalized histograms for $r_{u,3}$ calculated using the samples shown in Figure 6.13 from one run of the TMCMC algorithm when updating model class $\mathcal{M}_2$ with drift data.

constrained (Figure 6.12), but the second-story strength is somewhat less pinned down and the third-story strength has a relatively large uncertainty (Figure 6.13). This behavior can be explained by looking at Figure 6.10; since the first story is subjected to substantial yielding, much information about its ultimate strength is available, but since no appreciable yielding occurs in the third story, the data only provides information that low values of the third-story strength are not plausible, as can be seen in Figure 6.13 by the way the samples move between the prior and the posterior. Figure 6.14 also illustrates this by showing the evolution of the histogram of $r_{u,3}$. Note that for the higher values of ultimate story strength, the histogram for the final level has the same shape as the prior PDF. This is because there is very little information in the data about the upper limit of the ultimate strength of the third story, so the posterior PDF there is limited by the prior PDF. This makes sense, as the prior information must take precedence when no useful information is provided by the data.

The estimates of the stiffness parameters for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$ are still fairly tightly constrained, as can be seen from Table 6.2. However, the second and

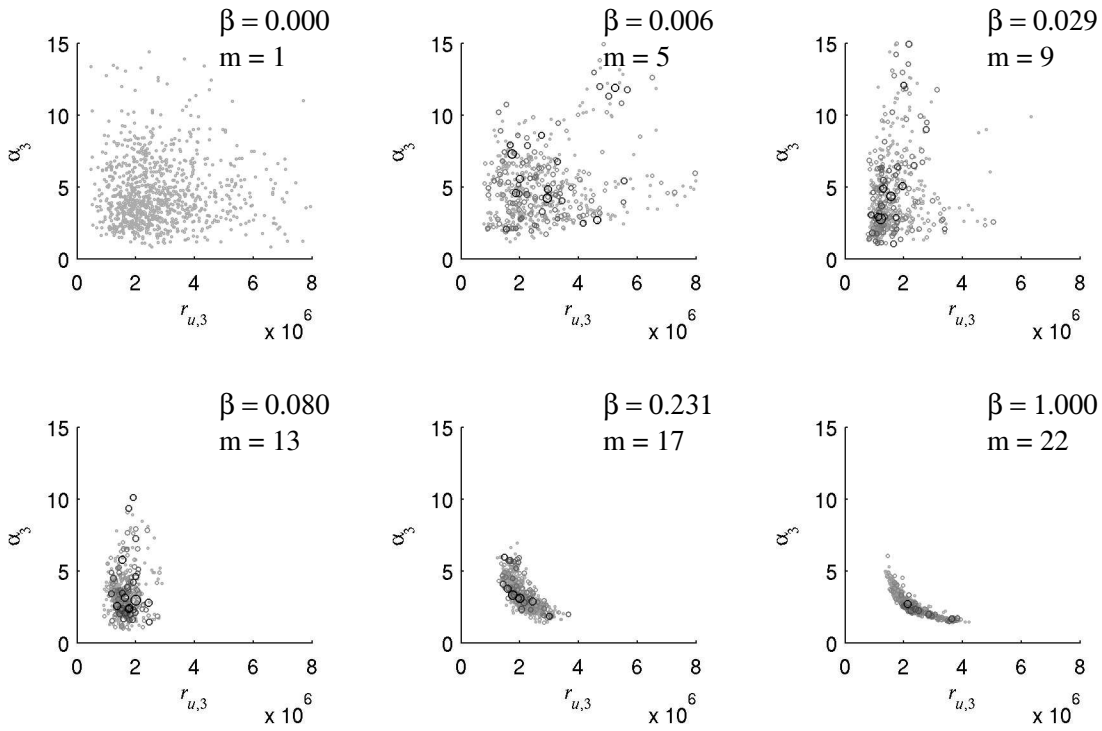


Figure 6.15: Plots of the samples in the $(r_{u,3}, \alpha_3)$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_3$ with drift data. Repeated samples are indicated by size and shading of the markers.

third stories exhibit more complicated behavior in the $(r_{u,i}, \alpha_i)$ space. For model classes $\mathcal{M}_1$ and $\mathcal{M}_2$, the manifold of optimal models is essentially constrained to move along a curve in the parameter space where only $r_{u,3}$ varies, because the value of α_i is pinned down for all three stories by the yielding in the first story. Model class $\mathcal{M}_3$, which does not constrain the elastic-to-plastic transition parameters to be equal for all stories, has a manifold of optimal models that exhibits interaction between the story ultimate strength and elastic-to-plastic transition parameters for the second and third stories. Figure 6.15 shows how samples in the $(r_{u,3}, \alpha_3)$ space converge during one run of the TMCMC algorithm when updating model class $\mathcal{M}_3$ with drift data. This type of geometry cannot be well-represented by an asymptotic approximation, and the numerical optimization necessary to make such an approximation is also difficult to perform without a good initial estimate of the parameters. The samples from updating model class $\mathcal{M}_4$ are also of interest. Figure 6.16 shows the aggregate samples for the final level of all three simulation runs using drift data, plotted in the $(r_{u,i}, \alpha_i)$

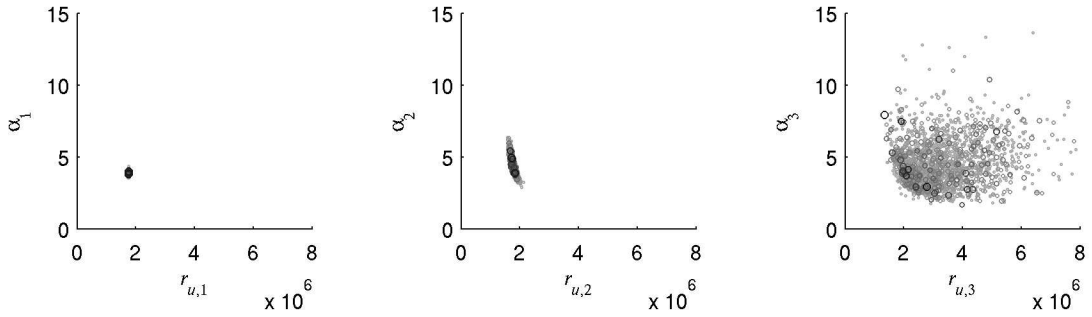


Figure 6.16: Aggregate samples for the final level of three runs of the TMCMC algorithm when updating model class $\mathcal{M}_4$ with drift data, plotted in the $(r_{u,i}, \alpha_i)$ space for all three stories.

space for each story. The first-story parameters are globally identifiable, as expected. The second-story parameters exhibit a manifold of optimal models similar to those shown in Figure 6.15 for model class $\mathcal{M}_3$. However, the third story parameters are not concentrated along a manifold, but are spread through the $(r_{u,3}, \alpha_3)$ space. A clear lower bound on the samples is evident, but the upper bounds are essentially constrained only by the prior PDF. Since the third story experienced almost no nonlinear behavior, this lower bound on the yielding parameters is all the information that can be extracted. It is possible that for model class $\mathcal{M}_3$, the third-story samples are concentrated along this lower bound (see Figure 6.15) in an attempt to compensate for the lack of viscous damping, because in this region the associated hysteretic energy dissipation is largest. In model class $\mathcal{M}_4$, the identified viscous-damping parameters are fairly close to the values used to generate the data, thus no additional energy dissipation is needed, and so the parameters associated with third-story yielding are essentially free away from the "lower bound" (Figure 6.16). This geometry of the posterior PDF cannot be represented with any type of asymptotic approximation and it is even more challenging to perform optimization here than for model class $\mathcal{M}_3$. Results of Bayesian model-class selection for model classes updated with drift data are shown in Table 6.3. The log-evidence and average log-likelihood function over the posterior PDF that are estimated from stochastic simulation are shown, along with the information gain, which is not directly estimated but rather calculated from

Model Class	Average Data Fit	Information Gain	Log Evidence	Probability
$\mathcal{M}_1$	8138.0	43.7	8094.3	0.0000
$\mathcal{M}_2$	8219.7	45.9	8173.8	0.8022
$\mathcal{M}_3$	8140.0	50.0	8090.0	0.0000
$\mathcal{M}_4$	8220.0	47.7	8172.4	0.1978

Table 6.3: Bayesian model-class selection results for drift data. The information gain is the difference between the average data fit and the log of the evidence, which are estimated using stochastic simulation

the other two quantities using Equation 2.16. The evidence clearly favors model class $\mathcal{M}_2$, which is unsurprising as it contains the model used to generate the data and has two fewer parameters than the other model class ($\mathcal{M}_4$) that contains the data-generation model. The improvement in the data-fit for model class $\mathcal{M}_2$ more than offsets the increased complexity (which is indicated by the larger information gain) compared to model class $\mathcal{M}_1$.

Model class $\mathcal{M}_3$ not only gives a poorer data-fit than $\mathcal{M}_2$ but it also has a larger information gain, while model class $\mathcal{M}_4$ exhibits a slightly improved data fit compared to $\mathcal{M}_2$ (possibly due to over-fitting of noise), which does not balance out the extra information needed to update the additional model parameters.

The results in Table 6.3 also show that model class $\mathcal{M}_3$ has a larger information gain than model class $\mathcal{M}_4$, which has two additional parameters. Figures 6.15 and 6.16 may explain this somewhat counter-intuitive result. While the samples for model class $\mathcal{M}_3$ are very tightly concentrated along a manifold in $(r_{u,3}, \alpha_3)$ space as seen in Figure 6.15, the samples for model class $\mathcal{M}_4$ are spread out through the same space, as seen in Figure 6.16 and as previously discussed. It seems reasonable that the extra information implied by the concentration of samples on the manifold might be greater than the information needed to identify the Rayleigh damping coefficients. This is an important point, as it shows that, unlike methods such as the AIC and BIC, the penalty against model complexity in Bayesian model-class selection is not based solely on the number of parameters, but rather on how much extra information is extracted from the data by the inclusion of these parameters.

6.2.3 Model Updating and Model-Class Selection with Acceleration Time Histories

The procedures applied to the drift response were also applied to the acceleration response. Three runs were performed using the TMCMC algorithm for each model class, each run using 1000 samples per level and needing between 19 to 23 levels to go from the prior PDF to the posterior PDF in a given run. Table 6.4 summarizes the results of the stochastic simulations for all four model classes and the most probable parameter values obtained using optimization. Again, the differences in the most probable parameter values obtained from simulation and optimization are less than one standard deviation.

In general, the parameter uncertainties are less than those for the case where drift data is used for updating, possibly because the acceleration records are richer in higher frequencies, so model behavior in the elastic-to-plastic transition region plays a more significant role in the response and constrains the corresponding model parameters more than for updating with drift records. Model class $\mathcal{M}_1$ exhibits a much lower identified value for the ultimate strength of the third story. This behavior is probably an attempt to compensate for the energy dissipated by the stiffness-proportional component of the viscous damping by using the hysteretic damping. The value may be lower than that obtained by updating with drift data because the offset that results from yielding, while important in the drift response, is at low frequencies and therefore is not so strongly represented in the acceleration response.

Of particular note is the identification of model class $\mathcal{M}_4$. Figure 6.17 shows plots of the aggregate samples plotted in the $(r_{u,i}, \alpha_i)$ space. Unlike the samples obtained by updating with drift records that are shown in Figure 6.16, the third-story yielding parameters are concentrated along a well-defined manifold in the $(r_{u,i}, \alpha_i)$ space, perhaps due to the increased importance of the high-frequency response in the acceleration records which provides more information for the updating.

Results for model-class selection using the acceleration data are shown in Table 6.5. Again, model class $\mathcal{M}_2$ is preferred, by an even wider margin than for updating with

Model Class	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$		$\mathcal{M}_4$	
	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.
K_1	2.503	2.501	2.500	2.500	2.523	2.526	2.500	2.499
10^8 N/m	(0.007)	[-0.002]	(0.007)	[0.000]	(0.011)	[0.003]	(0.007)	[-0.001]
K_2	2.533	2.533	2.495	2.493	2.528	2.529	2.499	2.498
10^8 N/m	(0.005)	[0.000]	(0.007)	[-0.002]	(0.006)	[-0.001]	(0.008)	[-0.001]
K_3	2.538	2.538	2.502	2.504	2.569	2.562	2.507	2.510
10^8 N/m	(0.006)	[0.000]	(0.008)	[0.002]	(0.028)	[-0.007]	(0.011)	[0.003]
$r_{u,1}$	1.808	1.806	1.747	1.747	1.814	1.814	1.746	1.746
10^6 N	(0.007)	[-0.002]	(0.007)	[0.000]	(0.006)	[0.000]	(0.024)	[0.000]
$r_{u,2}$	1.752	1.743	1.743	1.752	1.733	1.733	1.749	1.740
10^6 N	(0.019)	[-0.009]	(0.014)	[0.009]	(0.020)	[0.000]	(0.024)	[-0.009]
$r_{u,3}$	1.511	1.516	1.771	1.752	2.016	2.230	2.417	2.389
10^6 N	(0.020)	[0.005]	(0.064)	[-0.019]	(0.870)	[0.214]	(0.795)	[0.028]
α_1	3.873	3.922	4.023	4.032	3.771	3.749	4.056	4.068
	(0.087)	[0.049]	(0.075)	[0.009]	(0.080)	[-0.022]	(0.079)	[0.013]
α_2	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	4.075	4.050	3.931	4.015
	(fixed)	(fixed)	(fixed)	(fixed)	(0.193)	[-0.025]	(0.191)	[0.084]
α_3	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	1.826	2.014	2.456	2.690
	(fixed)	(fixed)	(fixed)	(fixed)	(0.814)	[0.188]	(0.735)	[0.234]
c_M	0.000	0.000	0.296	0.303	0.000	0.000	0.303	0.310
	(fixed)	(fixed)	(0.029)	[0.007]	(fixed)	(fixed)	(0.028)	[0.007]
c_K	0.000	0.000	2.801	2.764	0.000	0.000	2.788	2.721
10^{-4}	(fixed)	(fixed)	(0.159)	[0.037]	(fixed)	(fixed)	(0.167)	[-0.067]
σ^2_{acc}	0.496	0.493	0.271	0.270	0.482	0.472	0.274	0.269
$(\mathrm{m/s^2})^2$	(0.019)	[-0.003]	(0.010)	[-0.001]	(0.018)	[-0.010]	(0.010)	[-0.005]

Table 6.4: The estimated most probable parameter values and standard deviations for the marginal PDFs, calculated from samples, and for the joint PDF, calculated by optimization of the posterior, for model updating with acceleration data using both stochastic simulation and numerical optimization. Standard deviations are shown in parentheses, and the difference between the parameter values maximizing the marginal and joint PDFs is shown in brackets.

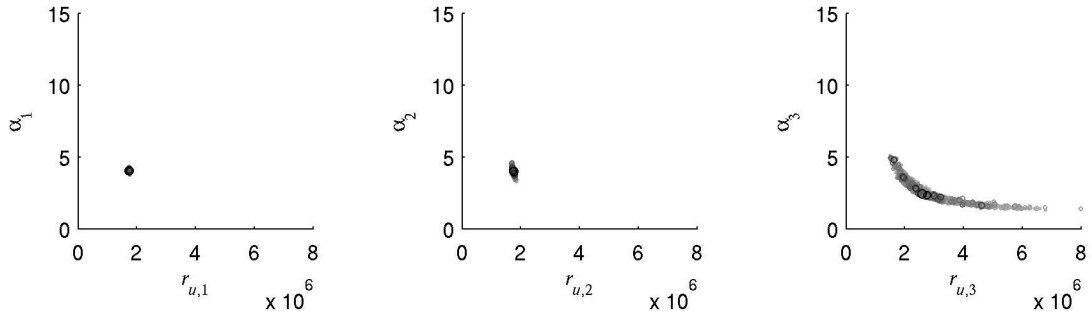


Figure 6.17: Aggregate samples for the final level of three runs of the TMCMC algorithm when updating model class $\mathcal{M}_4$ with acceleration data, plotted in the $(r_{u,i}, \alpha_i)$ space for all three stories.

Model Class	Average Data Fit	Information Gain	Log Evidence	Probability
$\mathcal{M}_1$	-1602.7	43.5	-1646.2	0.0000
$\mathcal{M}_2$	-1150.8	53.4	-1204.2	0.9989
$\mathcal{M}_3$	-1581.9	53.7	-1635.6	0.0000
$\mathcal{M}_4$	-1150.7	60.3	-1211.0	0.0011

Table 6.5: Bayesian model-class selection results for acceleration data. The information gain is the difference between the average data fit and the log of the evidence, which are estimated using stochastic simulation

drift data, since the improvement in data fit for model class $\mathcal{M}_4$ is negligible. It is also of some interest that model class $\mathcal{M}_3$ is preferred to model class $\mathcal{M}_1$, while the opposite was true when drift data was used for updating. The improvement in the data fit is large enough to justify the extra information extracted from the data.

6.2.4 Robust Predictive PDFs for Hysteretic Models

To illustrate how samples of hysteretic structural model parameters obtained through stochastic simulation may be used to incorporate uncertainty into predicted structural response, consider the following example.

Model class $\mathcal{M}_4$, presented in Section 6.2.1, is updated using acceleration data as before; however, in this case, only data from the first and third floors are available (no acceleration measurements at the second floor). Table 6.6 presents the estimated most probable values and standard deviations of the marginal posterior PDFs for each parameter, calculated from a single run of the TMCMC algorithm using 1000

	K_1 10^8 N/m	K_2 10^8 N/m	K_3 10^8 N/m	$r_{u,1}$ 10^6 N	$r_{u,2}$ 10^6 N	$r_{u,3}$ 10^6 N
Sim.	2.482 (0.019)	2.531 (0.025)	2.504 (0.013)	1.744 (0.011)	1.774 (0.039)	3.598 (1.232)
Opt.	2.478 [-0.004]	2.534 [0.003]	2.503 [-0.001]	1.744 [0.000]	1.761 [-0.013]	2.875 [-0.723]
	α_1	α_2	α_3	c_M	c_K 10^{-4}	σ^2 $(\mathrm{m/s^2})^2$
Sim.	3.96 (0.13)	3.74 (0.21)	4.21 (1.27)	0.307 (0.043)	2.758 (0.406)	0.248 (0.011)
Opt.	3.96 [0.00]	3.77 [0.03]	4.69 [0.48]	0.319 [0.012]	2.597 [-0.161]	0.243 [-0.005]

Table 6.6: The estimated most probable parameter values and standard deviations for the marginal PDFs, calculated from samples, and for the joint PDF, calculated by optimization of the posterior, for model updating with partial acceleration data using both stochastic simulation and numerical optimization. Standard deviations are shown in parentheses, and the difference between the parameter values maximizing the marginal and joint PDFs are shown in brackets.

samples per level. Also presented are estimates of the most probable parameter values obtained by optimization of the joint posterior PDF. Samples are plotted on the $(r_{u,i}, \alpha_i)$ sub-spaces in Figure 6.18. As might be expected, there are large uncertainties in the estimates for parameters associated with the third-story yielding inelastic response.

The data-generating system (specified in Section 6.2) is then subjected to a different input excitation, specifically the ground motion obtained at the Rinaldi Receiving Station during the 1994 Northridge earthquake. This record features somewhat larger

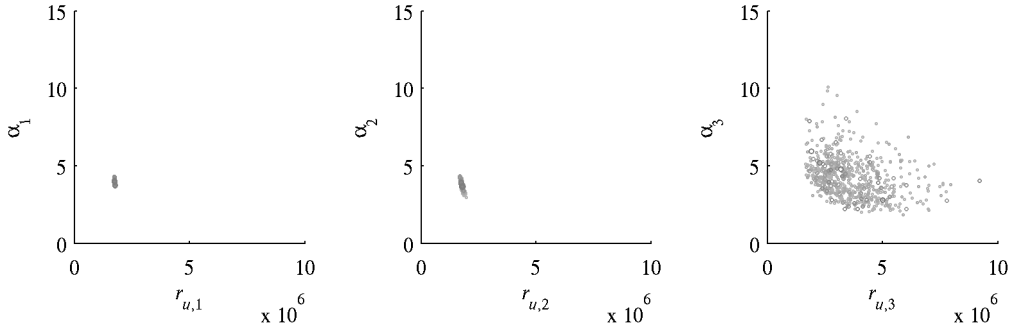


Figure 6.18: Samples for model updating with partial acceleration data (first and third floors), plotted in the $(r_{u,i}, \alpha_i)$ space for all three stories.

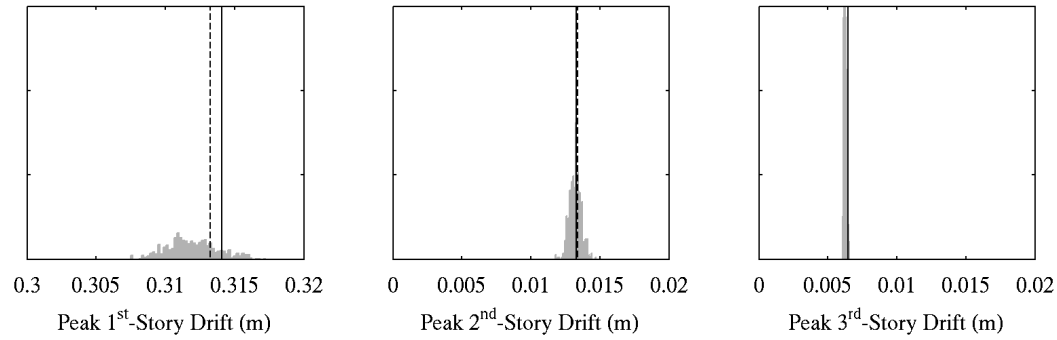


Figure 6.19: Predicted maximum inter-story drifts for the updated model class in response to the Rinaldi Receiving Station record. The shaded area is the normalized histogram of the maximum predicted drift responses generated using samples obtained from stochastic simulation. The dashed line represents the predicted maximum drift based on parameter values obtained by optimization, and the solid line indicates the actual response of the data-generating system.

accelerations compared to the Olive View hospital ground motion. Optimal and robust predictive PDFs for the peak inter-story drifts are calculated from Equations 2.7 and 3.9, respectively. Comparisons between the robust PDF, the optimal PDF, and the actual values of the peak drifts exhibited by the data-generating system are shown for each story in Figure 6.19.

Note that the prediction for maximum first-story drift has the greatest uncertainty. This may seem surprising at first, as there is much less uncertainty associated with the identified first-story model parameters compared to those for the third story. However, the first story is where the greatest non-linear response occurs, and therefore the first-story drifts are much more sensitive to the total uncertainty in the predictive model than the much smaller third-story drifts.

6.3 Deteriorating Hysteretic Models

An important consideration in modeling structural responses to very strong shaking is that damage to structural and non-structural elements can significantly alter the dynamic characteristics of the structure during the shaking. Identifying time-varying structural parameters to capture this deterioration in structural behavior is difficult

and limits the applicability of the model for prediction of future response. However, the class of Masing models can be straightforwardly extended to the case of deteriorating strength and stiffness using time-invariant parameters. As might be expected, the issues with ill-conditioning and non-identifiability present for non-deteriorating models persist, and even worsen, making stochastic simulation a highly relevant tool for studying such systems.

6.3.1 Deteriorating Masing Models

The basis of the deteriorating Masing models is again the DEM, specifically the displacement-controlled deteriorating DEM introduced by Iwan and Cifuentes (1986). The model is similar to that described in the previous section; however, in this model the individual elasto-plastic elements are allowed to "break" if the displacement exceeds a threshhold value, which was defined as $\mu x_{y,i}$, where $x_{y,i}$ is the yield displacement for the ith element and μ is the *breaking ductility ratio*, which for simplicity is assumed to be the same for all elasto-plastic elements. Cifuentes and Iwan (1989) successfully performed identification of this model using data from actual structures. Ashrafi et al. (2005) identified damage using a class of infinite deteriorating DEMs, but only for single-degree-of-freedom cases where the displacement and restoring force time histories are available.

A generalized class of deteriorating Masing models have been constructed (Chiang, 1992, 1999). This was accomplished by studying the integral form of the restoring force for DEMs, given in Equation 6.2. Chiang added a damage index function, $d(x, r^*) \in [0, 1]$, such that the restoring force during initial loading is then given by

$$r(x) = \int_0^{kx} r^* \phi(r^*) d(x, r^*) dr^* + kx \int_{kx}^{\infty} \phi(r^*) dr^* \tag{6.18}$$

where $\phi(r^*)d(x, r^*)$ and $1 - \phi(r^*)d(x, r^*)$ represent the fraction of undamaged and damaged elements, respectively, with yield strengths in the interval $(r^*, r^* + dr^*)$. Chiang determined the integrals giving the restoring forces for the unloading and reloading cases.

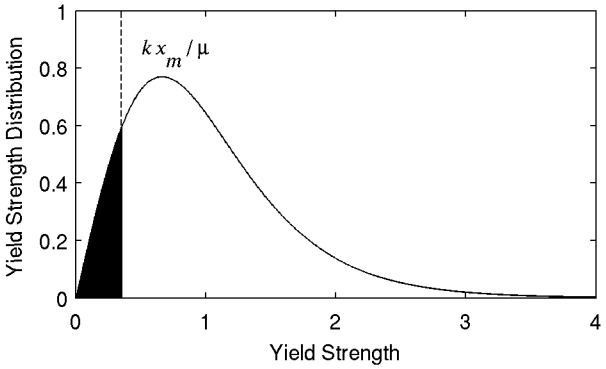


Figure 6.20: Yield-strength distribution for a deteriorating hysteretic system. The shaded portion indicates elements that have broken after having exceeded the prescribed force level indicated by the dashed line.

It was found that the behavior of the displacement-controlled deteriorating DEM of Iwan and Cifuentes could be reproduced by choosing the damage index function:

$$d(x, r^*) = 1 - H\left(\frac{kx_m}{\mu} - r^*\right) \tag{6.19}$$

where $H(.)$ is the Heaviside step function, μ is the previously discussed breaking ductility ratio, and x_m is the maximum experienced displacement, defined as

$$x_m(t) = \max_{\tau \leq t} |x(\tau)| \tag{6.20}$$

This can be viewed as all elements with yield strengths below the level kx_m/μ having been broken and no longer contributing to the restoring force, as illustrated in Figure 6.20. Chiang (1992) developed expressions for the unloading and reloading branch behavior for a hysteretic system with an initial loading curve as defined in Equation 6.3, and showed that they could be combined with the extended Masing rules, discussed in Section 6.1, to give deteriorating Masing models. Figure 6.21 shows the initial loading curves for deteriorating Masing models (based on a generalized Rayleigh distribution with $\eta = 1$ for the yield strength) with several different values of the breaking ductility ratio μ. The damage-index function given in Equation 6.19 also leads to the following expression for the fraction of broken elements,

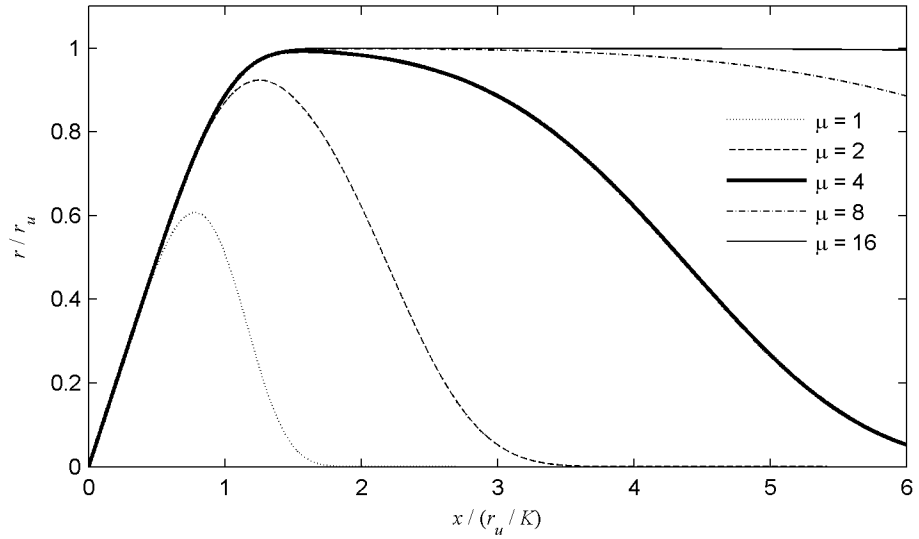


Figure 6.21: Initial loading curves for a deteriorating Masing model with a generalized Rayleigh yield-strength distribution (with $\eta = 1$), plotted for different values of the breaking ductility ratio μ

denoted as the damage fraction $D(t)$:

$$D(t) = \int_0^{kx_m/\mu} \phi(r^*)dr^* \tag{6.21}$$

which is simply the area of the shaded region in Figure 6.20. This might be considered a sort of generalized damage index, and by examining the restoring force in Equations 6.2 and 6.18, it can be considered to give an expression for the fractional loss in small-amplitude stiffness $D_{stiff}(t)$. The expression for the fractional loss in ultimate strength is given by

$$D_{str}(t) = \int_0^{kx_m/\mu} r^*\phi(r^*)dr* \tag{6.22}$$

Note that if the element is based on a generalized Rayeigh distribution for the element yield strength, as defined by Equations 6.10 through 6.13, then $D_{stiff}(t)$ and $D_{str}(t)$

are given by the closed-form expressions

$$D_{stiff}(t) = 1 - \exp\left[-\nu\left(\frac{x_m}{\mu}\right)^{\eta}\right] \quad (6.23)$$

$$D_{str}(t) = \Gamma\left(\nu\left(\frac{x_m}{\mu}\right)^{\eta}, \frac{\eta+1}{\eta}\right) \quad (6.24)$$

where the constant ν and the complete and incomplete gamma functions, $\Gamma(.)$ and $\Gamma(.,.)$, are defined as for Equation 6.13.

Note that there may be significant differences in stiffness and strength losses. For example, the deteriorating system shown in Figure 6.20 has about a 9% loss of stiffness, but only around a 2% loss in strength, since the broken elements have relatively low strengths, and thus contribute only a small amount to the total stength. Recall that for the Distributed Element Model, which can be viewed as the physical basis for the Masing hysteretic model, each differential element contributes equally to the elastic stiffness (see Figure 6.2), so the loss of lower-strength elements has a greater effect on the small-amplitude stiffness of the system.

6.4 Studies of Deteriorating Systems with Simulated Data

Data is generated using a three-story Masing shear-building model excited by the Sylmar ground motion, as described in Section 6.2. However, instead of using the Bouc-Wen-like initial loading curve described by Equation 6.9, the relationship between inter-story drift and inter-story shear force is described by Equation 6.11. This is done because the rules defining the deteriorating hysteresis loops are more easily implemented. The value of the elastic-to-plastic transition parameter is set to $\eta_i = 3$, compared to $\alpha_i = 4$ for the Masing shear model used in Section 6.2. Comparison of the curves in Figures 6.4 and 6.7 shows that the two models should produce relatively similar initial loading curves in the non-deteriorating case. Values for the floor masses, the small-amplitude story stiffnesses, ultimate story strengths, and viscous-

damping matrix are identical to those used in Section 6.2. Two types of structures are considered: a non-deteriorating system and a deteriorating system with a breaking ductility ratio of $\mu_i = 6$ for each story.

Simulated acceleration data is generated for each system, 500 data points for each story at a sampling interval of 0.02 seconds. Gaussian discrete white noise is added to the responses, with a standard deviation of 0.5 $\mathrm{m/s^2}$, which is equal to about 20% of the root-mean-square value of the time histories. The data set corresponding to the non-deteriorating structure will be henceforth referred to as $\mathcal{D}_{yield}$ and the set corresponding to the deteriorating system as $\mathcal{D}_{dmg}$.

The identification model class is a deteriorating hysteretic model. The likelihood function is given in Equation 6.17. The exact viscous-damping matrix was used, since the identifiability of this matrix had already been explored in the previous example. Similarly, the identification model constrained the values of the elastic-to-plastic transition parameter η_i to be equal at each story. The breaking ductility ratio μ_i is similarly constrained. The resulting identification model has nine uncertain parameters, which are listed in the first row of Table 6.7.

Prior PDFs for small-amplitude stiffness, story strength, and prediction-error variance are unchanged from the previous example. The prior PDFs for η and μ are lognormal, with logarithmic means of ln(3) and ln(10), respectively, and logarithmic standard deviations of 0.5.

6.4.1 Model Updating with Data from a Non-Deteriorating System

Updating was performed with the data $\mathcal{D}_{yield}$ using TMCMC with 500 samples per level. Three runs were performed, with 20 to 21 levels needed to converge from the prior to the posterior PDF. The results are summarized in Table 6.7, which lists the most probable values for the marginal PDF of each parameter, calculated from the aggregate samples of the three runs. For comparison, the most probable parameter values calculated by numerical optimization of the posterior PDF are also given. Note

	K_1 10^8 N/m	K_2 10^8 N/m	K_3 10^8 N/m	$r_{u,1}$ 10^6 N	$r_{u,2}$ 10^6 N	$r_{u,3}$ 10^6 N	η	μ	σ^2 $(\mathrm{m/s^2})^2$
Sim.	2.497 (0.013)	2.501 (0.020)	2.497 (0.013)	1.747 (0.007)	1.757 (0.027)	1.633 (0.144)	3.02 (0.06)	20.24 (5.58)	0.280 (0.010)
Opt.	2.486 [-0.011]	2.507 [0.006]	2.504 [0.007]	1.743 [-0.004]	1.741 [-0.016]	1.558 [-0.075]	3.08 [0.06]	21.60 [1.37]	0.275 [-0.005]

Table 6.7: Most probable parameter values for marginal PDFs updated using data from the non-deteriorating system and most probable parameter values for the joint PDF obtained by optimization of the posterior. Standard deviations for simulation results are shown in parentheses. Differences between the optimal results for the joint PDF and the marginal PDFs are shown in brackets.

	Final Stiffness Loss	Final Strength Loss
1st Story	0.0042 (0.0032)	0.0006 (0.0006)
2nd Story	0.0002 (0.0002)	0.0000 (0.0000)
3rd Story	0.0000 (0.0000)	0.0000 (0.0000)

Table 6.8: Mean values of final fractional stiffness and strength loss for updating using data from the non-deteriorating system. Standard deviations for simulation results are shown in parentheses.

that the initial estimates of the parameters for the optimization were based on the results of the stochastic simulation, as the optimization was very sensitive to initial conditions.

Figure 6.22 shows how the samples converge in the $(r_{u,3}, \mu)$-space. Samples for most of the parameters are fairly well pinned down. However, the samples for μ are widely spread. Since there is no actual damage, instead of identifying a particular value for the breaking ductility ratio, the stochastic simulation results indicate a minimum boundary for μ as demonstrated in Figure 6.22.

The fractional loss of strength and stiffness for any given sample can be calculated from the predicted structural response using Equation 6.21. The final damage fractions (damage at the end of the response record) are calculated for each sample, with the mean and standard deviation given in Table 6.8. These results are close to the exact values, which are zero because there is no deterioration in the hysteretic system used to generate the data.

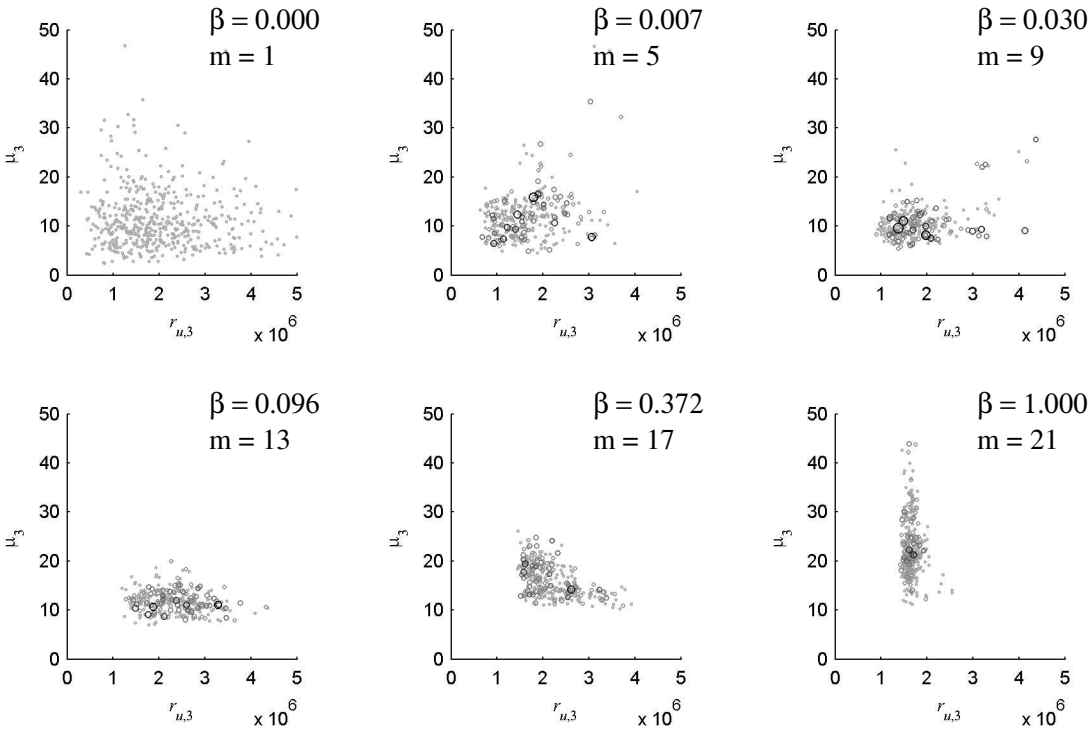


Figure 6.22: Plots of the samples in the $(r_{u,3}, \mu)$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_3$ with acceleration data generated with a non-deteriorating system. Repeated samples are indicated by size and shading of the markers.

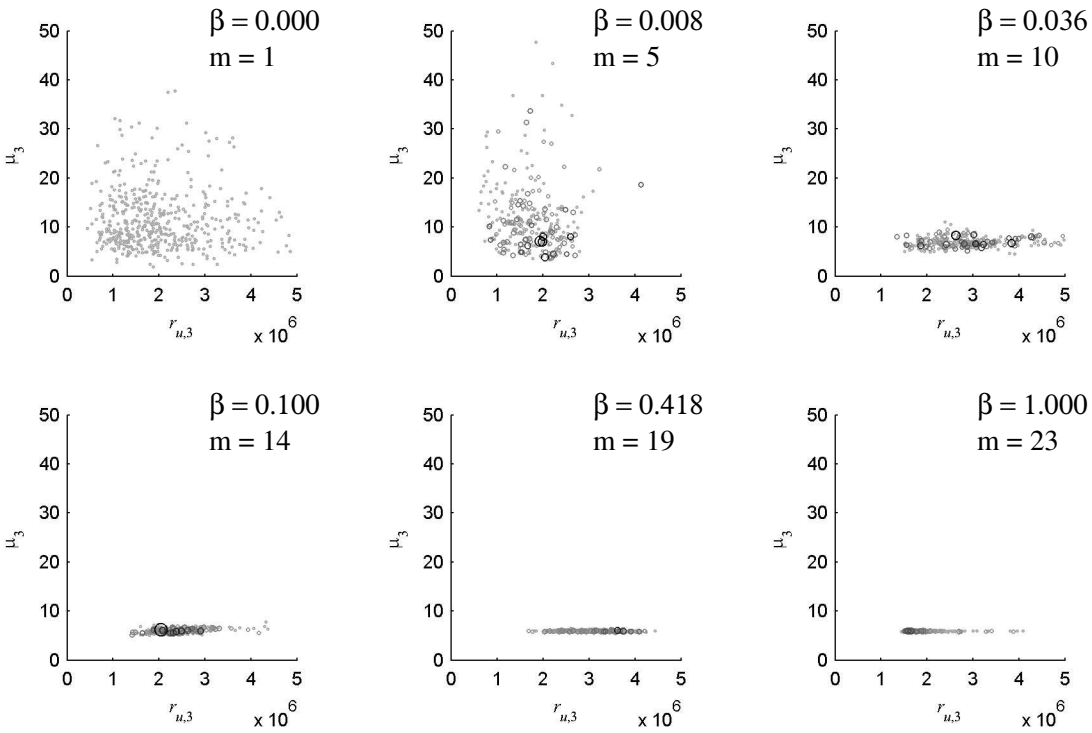


Figure 6.23: Plots of the samples in the $(r_{u,3}, \mu)$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_3$ with acceleration data generated with a deteriorating system. Repeated samples are indicated by size and shading of the markers.

6.4.2 Model Updating with Data from a Deteriorating System

Updating with the data $\mathcal{D}_{dmg}$ using TMCMC was performed using 500 samples per level for three runs, with 23 to 24 levels needed to converge from the prior to the posterior PDF. Figure 6.23 shows how samples converge in the $(r_{u,3}, \mu)$-space for one run of the TMCMC algorithm. The samples for the final level are concentrated into a much smaller region compared to Figure 6.22, which shows the samples generated by updating with $\mathcal{D}_{yield}$. Results for the most probable values and standard deviations, obtained from the samples, are summarized in Table 6.9, along with the most probable parameter values of the joint PDF identified by numerical optimization of the posterior PDF. Generally, the identified values are close to the values of the parameters in the model used to generate the data.

	K_1 10^8 N/m	K_2 10^8 N/m	K_3 10^8 N/m	$r_{u,1}$ 10^6 N	$r_{u,2}$ 10^6 N	$r_{u,3}$ 10^6 N	η	μ	σ^2 m/s^2
Sim.	2.487 (0.015)	2.509 (0.024)	2.512 (0.015)	1.756 (0.006)	1.740 (0.022)	1.604 (0.307)	3.04 (0.06)	5.86 (0.07)	0.255 (0.010)
Opt.	2.488 [0.001]	2.510 [0.001]	2.520 [0.008]	1.757 [0.001]	1.738 [-0.002]	1.555 [-0.049]	3.06 [0.02]	5.82 [-0.04]	0.238 [-0.005]

Table 6.9: Most probable parameter values for marginal PDFs updated using data from the deteriorating system and most probable parameter values for the joint PDF obtained by optimization of the posterior. Standard deviations for simulation results are shown in parentheses. Differences between the optimal results for the joint PDF and the marginal PDFs are shown in brackets.

Story	Final Stiffness Loss	Final Strength Loss
1st	0.1371 (0.0033)	0.0609 (0.0016)
2nd	0.0053 (0.0005)	0.0008 (0.0001)
3rd	0.0013 (0.0005)	0.0001 (0.0001)

Table 6.10: Mean values of final fractional stiffness and strength loss for updating using data from the deteriorating system. Standard deviations for simulation results are shown in parentheses.

As shown previously, the damage fraction may be calculated from Equation 6.23. The mean and standard deviation of the final maximum damage fractions calculated for the samples are given Table 6.10. The evolution of the damage fraction in time may also be tracked. Figure 6.24(a) shows the predicted inter-story drift time histories calculated using the most probable parameter values given in Table 6.9, which are then used to calculate the time histories of the story damage fractions, which are shown in Figure 6.24(b).

The effect of the deterioration on the small-amplitude stiffness of the structure can be directly confirmed by examining the predicted hysteresis loops. Figure 6.25 shows the hysteresis loop for the first story of the deteriorating structures predicted using the parameter values from the sample with the largest value of log-likelihood (and therefore expected to be drawn from a high-probability region of the posterior PDF). The maximum inter-story drift occurs at point A, indicating the moment when the maximum damage is achieved. Initial reloading from point A has an apparent

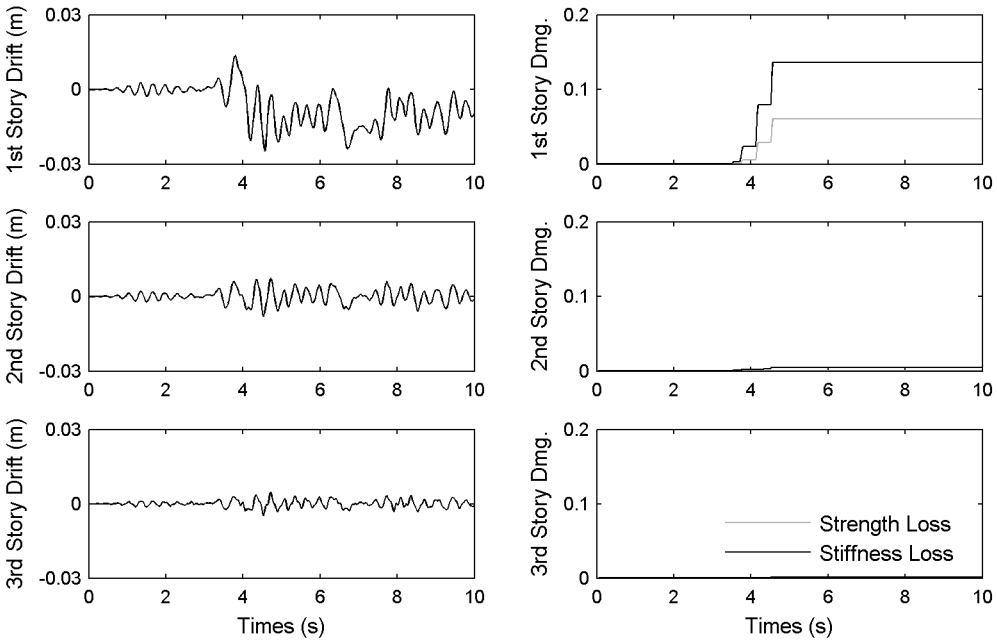


Figure 6.24: Time histories of (a) inter-story drifts and (b) story damage fractions predicted using the set of most probable parameters updated with $\mathcal{D}_{dmg}$.

stiffness that is 86% of the initial small-amplitude stiffness, corresponding with the identified damage fraction of 14% given in Table 6.10. There is no similar method for directly confirming the strength loss, as the ultimate story strength is not approached in the portion of the record after the point of maximum experienced drift.

6.5 Preliminary Study with Real Data from a Seven-story Building Damaged in an Earthquake

The building selected for this study is a seven-story hotel located in Van Nuys, California, in the San Fernando Valley of Los Angeles County, which will henceforth be referred to as the Van Nuys hotel. The Van Nuys hotel is a reinforced-concrete moment frame building constructed in 1966. The building was lightly damaged in the 1971 San Fernando Earthquake, and severely damaged in the 1994 Northridge earthquake, which has made it the focus of numerous studies, e.g., Jennings (1971); Scholl

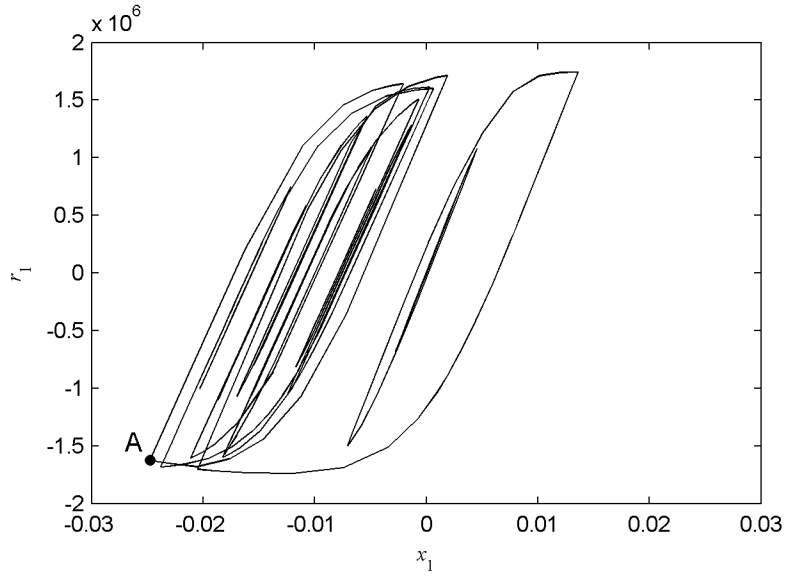


Figure 6.25: Hysteresis loop for the first-story response of the deteriorating model. Point "A" indicates the maximum displacement, corresponding to peak damage.

et al. (1982); Islam (1996); Li and Jirsa (1998); Beck et al. (2002); Ching et al. (2004). Trifunac et al. (1999) provides detailed descriptions and photographs of damage to the structure resulting from the Northridge earthquake.

The Van Nuys hotel was instrumented with sixteen accelerometers during the 1994 Northridge earthquake; of particular interest are measurements of the east-west accelerations at five floors, as the south frame of the building (which provides lateral stiffness in the east-west direction) was heavily damaged during the event. Accelerometers oriented in the east-west direction were installed in the southeast corner of the ground floor and near the east wall of the structure on the second, third, and sixth floors and the roof. Unfortunately, most of the damage to the south frame occurred at the beam-column connections at the top of the fourth story (Trifunac et al., 1999), as shown in Figures 6.26 and 6.27. This damage would be most clearly reflected in the unmeasured fifth-floor accelerations.

The Northridge earthquake records for the east-west direction are used here for system identification of the Van Nuys hotel. The ground floor accelerations are used as the input ground motion, so only four channels of output data are available, making this an extremely ill-conditioned system-identification problem when using the

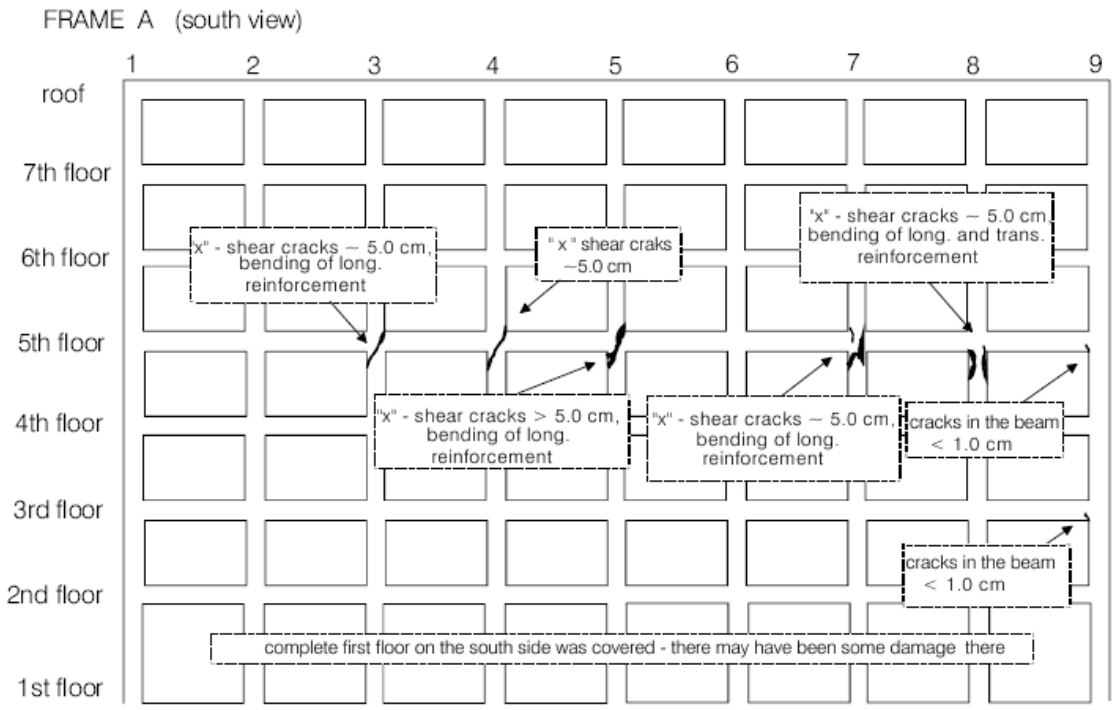


Figure 6.26: Schematic representation of damage to the south frame of the building from the Northridge earthquake. Taken from Trifunac et al. (1999).

deteriorating Masing hysteretic model class.

6.5.1 Identification Model

Four model classes are considered, which are all seven-story deteriorating Masing shear-building models based on a generalized Rayleigh distribution for the yield-strength, similar to the three-story model used in Section 6.4. The input forces are computed using the ground floor accelerations as the input ground motion and assuming a rigid foundation.

The mass matrix for the structure is diagonal, with values for the floor masses, taken from Ching et al. (2004), of 1.2002×10^5 kg for the first floor, 1.2002×10^5 kg for the second through seventh floors, and 0.9587×10^5 kg for the roof. As in the other studies presented in this chapter, the masses are assumed to be known accurately and thus the mass matrix is held constant.

Prior PDFs for the initial story stiffnesses are taken as independent lognormal

Figure 6.27: Photograph of shear cracks occuring in the south frame of the structure during the Northridge earthquake. Taken from Trifunac et al. (1999).

distributions, with the logarithmic means for each PDF given by the log of the values for story stiffness used in a seven-story linear-shear-building model presented in Ching et al. (2004). These values for the small-amplitude stiffness are 1.00×10^8 N/m for the first story, 1.20×10^8 N/m for the second story, 0.98×10^8 N/m for the third story, and 0.80×10^8 N/m for remaining four stories. The logarithmic standard deviation for each PDF is 0.5. The prior PDFs for story strength are also independent lognormals, with logarithmic means equal to the above values for small-amplitude stiffness multiplied by 0.01 m, and logarithmic standard deviations of 0.5.

The prior PDFs for the elastic-to-plastic transition parameter η_i and the breaking ductility ratio μ_i of each story are lognormal with means of 2 and 10, and logarithmic standard deviations of 1 and 0.5, respectively.

A stiffness-proportional viscous-damping matrix is included to provide small-amplitude damping, with the prior PDF for the constant of proportionality c_K given by a uniform distribution from 0 to 0.02. The stiffness-proportional damping matrix is chosen because it most closely conforms to the viscous-damping matrix used in Ching et al. (2004).

An inverse gamma prior is used for the prediction-error variance, with the shape parameter $\alpha = 3$ and the scaling parameter $\beta = 1.87$, the average of the variance in each of the four output acceleration records.

For model class $\mathcal{M}_1$, the elastic-to-plastic transition parameters and the breaking ductility ratios are constrained to be the same for each story; that is, $\eta_i = \eta$ and $\mu_i = \mu$, $i = 1, \ldots, 7$. In model class $\mathcal{M}_2$, the elastic-to-plastic transtion parameter is allowed to vary between stories with the breaking ductility ratio still constrained, while for $\mathcal{M}_3$, the breaking ductility ratios are freed while the elastic-to-plastic transition parameters are kept constrained. Model class $\mathcal{M}_4$ allows both the elastic-to-plastic transition parameter and breaking ductility ratio to vary between stories. Model class $\mathcal{M}_1$ has a total of 18 uncertain parameters, while model classes $\mathcal{M}_2$ and $\mathcal{M}_3$ each have 24 uncertain parameters and model class $\mathcal{M}_4$ has 30 uncertain parameters. All of these model classes therefore have a relatively large number of uncertain parameters, much more than would normally be used in an ill-conditioned system-

Model Class	$K_i,\ i = 1, \ldots, 7$	$r_{u,i},\ i = 1, \ldots, 7$	$\eta_i,\ i = 1, \ldots, 7$	$\mu_i,\ i = 1, \ldots, 7$	c_K	σ
$\mathcal{M}_1$	free	free	$= \eta_1$	$= \mu_1$	free	free
$\mathcal{M}_2$	free	free	free	$= \mu_1$	free	free
$\mathcal{M}_3$	free	free	$= \eta_1$	free	free	free
$\mathcal{M}_4$	free	free	free	free	free	free

Table 6.11: Summary of parameters for candidate model classes.

identification problem. The candidate model classes are summarized in Table 6.11.

6.5.2 Model Updating Using a Hybrid TMCMC-Gibbs Sampler Algorithm

A very recent suggestion by Ching (2006) to improve the performance of the TMCMC algorithm involves combining the method with the Gibbs sampler. In this case, updating for all parameters except the prediction-error variance is carried out as before, with the prediction-error variances fixed at the values from the previous level. The resulting samples may then be considered to be distributed from a conditional intermediate PDF. Then, for each sample of the parameter values, the conditional distribution for the prediction-error variances is sampled. This method requires that a conjugate prior be chosen for the prediction-error variance to allow for fast sampling, as explained in Section 4.1.1, and hence the choice of the previously mentioned inverse gamma prior.

The hybrid TMCMC-GS algorithm was used to generate Markov chains for each of the three model classes, with 500 samples per level for model classes $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$, and 2000 samples per level for $\mathcal{M}_4$, due to the larger number of parameters. 26 to 28 levels were needed to converge from the prior to the posterior PDF. Figure 6.28 shows the projection of the samples generated using TMCMC-GS onto the $(K_i, r_{u,i})$ sub-spaces for each story.

Figure 6.29 shows the predicted floor accelerations for the models in each model class that are given by the sample with the largest value for the posterior, which can reasonably be assumed to be close to the most probable parameter values in the model class, plotted with the available acceleration data. These maximum posterior samples

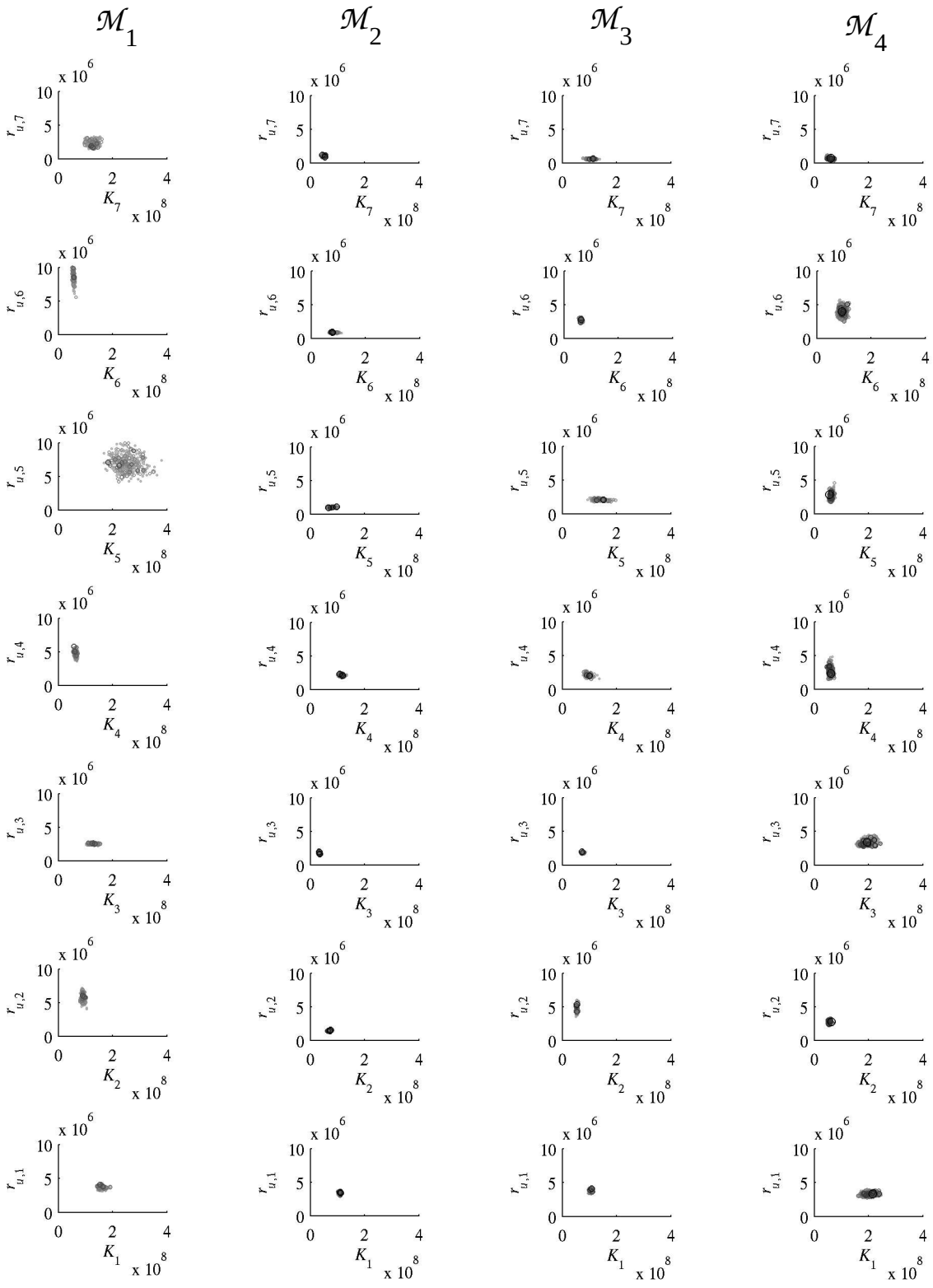


Figure 6.28: Posterior samples based on the Northridge earthquake data for each model class projected onto the $(K_i, r_{u,i})$ subspace for each story.

differ quite substantially between model classes, as implied by Figure 6.28, because the model classes are nearly unidentifiable. Note that while the model predictions are able to match the lower-frequency behavior of the system; the relatively simple shear-building identification model cannot generate the higher-frequency oscillations present in the data. As expected, the model predictions for acceleration are fairly close at the floors where acceleration measurements are available (second, third, and sixth floor and roof), but differ somewhat for unobserved floors. Figures 6.30 through 6.33 show hysteresis loops for the first four stories that are predicted by the same models. While the largest inter-story drifts are predicted for the third story, model class $\mathcal{M}_2$ also predicts significant non-linear behavior in the fourth story, which is consistent with the observed damage. Model classes $\mathcal{M}_1$, $\mathcal{M}_3$ and $\mathcal{M}_4$ indicate nearly linear behavior at the fourth story.

The final fractional loss in stiffness and strength is calculated for each sample using Equations 6.23 and 6.24. The mean and standard deviations of the final damage fractions are shown for each model class in Table 6.12. These results show more clearly than the hysteresis loops, that significant (nearly 50%) loss of strength is identified at the fourth story for model class $\mathcal{M}_2$, but not for the other three model classes (less than 2% loss of strength).

While these results are interesting in that the predicted behavior for model class $\mathcal{M}_2$ conforms to observed damage in the structure, further investigation is needed. Based on the results of this study, perhaps a larger number of samples than used here will be needed to ensure that the posterior PDF is better represented. With a larger number of samples, model-class selection could also be performed.

Another issue to consider is the effect of the viscous-damping matrix. Though the first-mode critical damping ratios for all model classes seem reasonable (ranging from 2 to 4%), the forces associated with viscous damping are unrealistically large, an issue recently addressed by Hall (2006). Figure 6.34 shows that the damping force is nearly equal in magnitude to the restoring force. Updating of model classes that omitted the viscous-damping matrix completely were conducted, however, the updated models predicted substantial yielding in the upper four stories. Non-linear behavior occurred

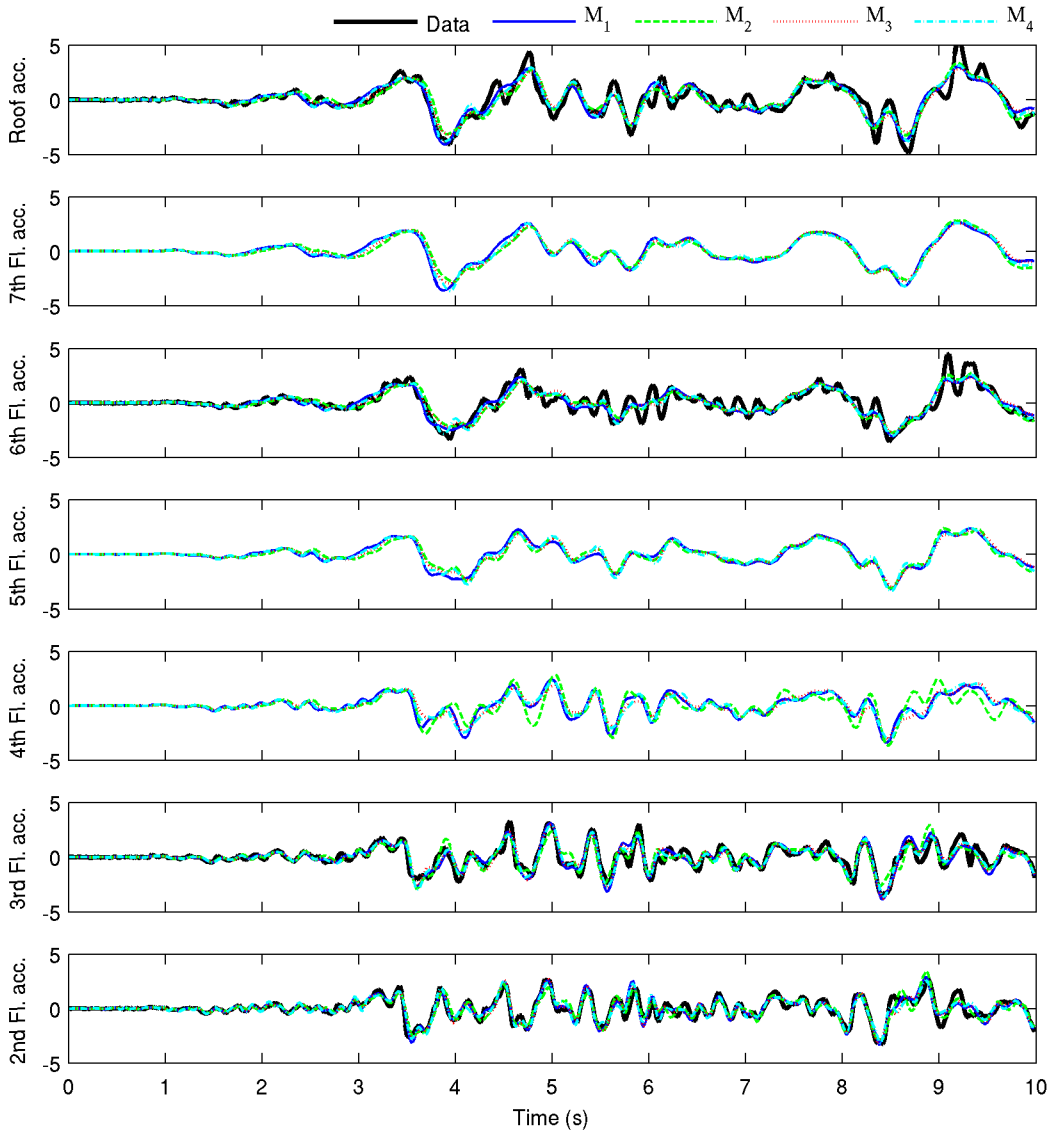


Figure 6.29: Predicted floor accelerations for models corresponding to the maximum posterior samples for each of the four model classes, plotted with the available data from the Northridge earthquake.

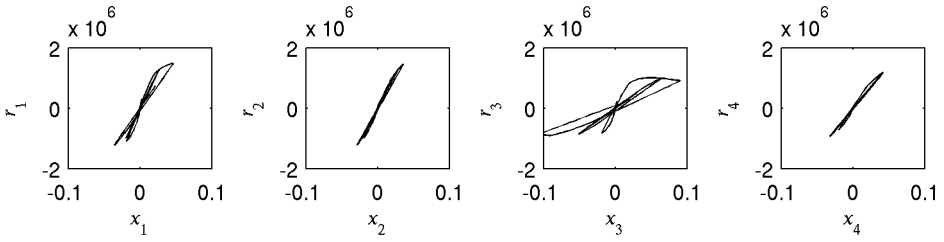


Figure 6.30: Hysteresis loops for the Northridge earthquake response of the first four stories of a model corresponding to the maximum posterior sample from model class $\mathcal{M}_1$.

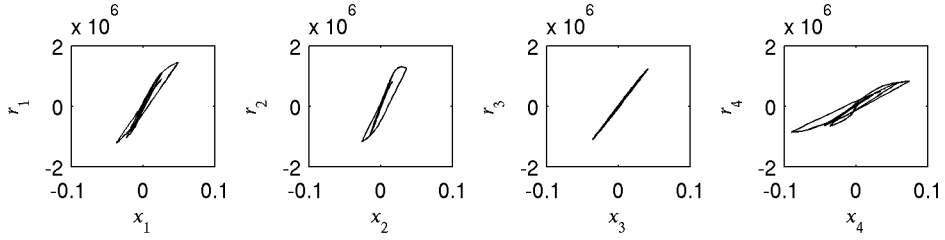


Figure 6.31: Hysteresis loops for the Northridge earthquake response of the first four stories of a model corresponding to the maximum posterior sample from model class $\mathcal{M}_2$.

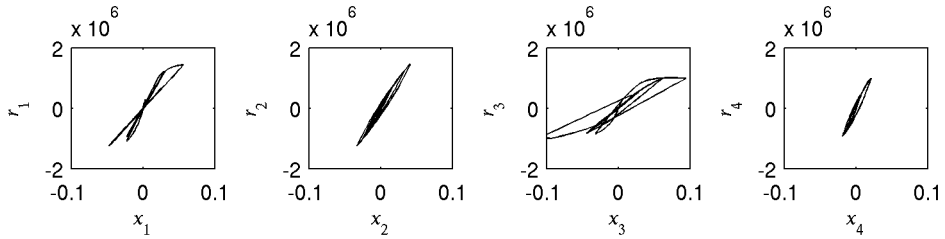


Figure 6.32: Hysteresis loops for the Northridge earthquake response of the first four stories of a model corresponding to the maximum posterior sample from model class $\mathcal{M}_3$.

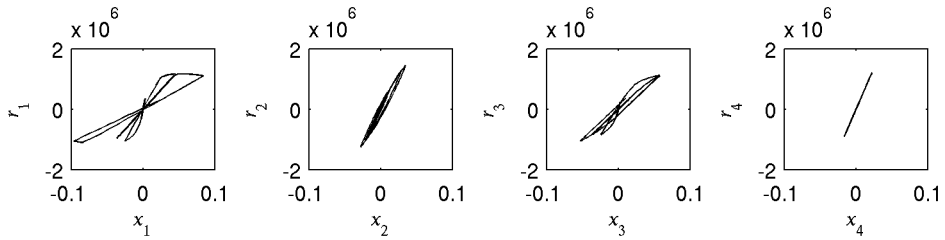


Figure 6.33: Hysteresis loops for the Northridge earthquake response of the first four stories of a model corresponding to the maximum posterior sample from model class $\mathcal{M}_4$.

	Final Stiffness Loss				Final Strength Loss			
	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
1st Story	0.8743 (0.0118)	0.4437 (0.0245)	0.7676 (0.0119)	0.8645 (0.0106)	0.2833 (0.0111)	0.0987 (0.0061)	0.2693 (0.0192)	0.2972 (0.0191)
2nd Story	0.4926 (0.0383)	0.4707 (0.0243)	0.1446 (0.0149)	0.0786 (0.0038)	0.0200 (0.0364)	0.0644 (0.0066)	0.0021 (0.0005)	0.0047 (0.0030)
3rd Story	0.9220 (0.0089)	0.7052 (0.0296)	0.8431 (0.0115)	0.9391 (0.0043)	0.4049 (0.0184)	0.2232 (0.0199)	0.3873 (0.0180)	0.3339 (0.0222)
4th Story	0.4799 (0.0212)	0.9219 (0.0106)	0.2416 (0.0231)	0.000 (0.000)	0.0184 (0.0055)	0.4914 (0.0262)	0.0085 (0.0023)	0.000 (0.000)
5th Story	0.6897 (0.0281)	0.000 (0.0000)	0.1523 (0.0177)	0.0079 (0.0168)	0.0828 (0.0118)	0.000 (0.0000)	0.0024 (0.0007)	0.0004 (0.0010)
6th Story	0.4293 (0.0233)	0.1650 (0.0304)	0.3651 (0.0258)	0.5736 (0.0359)	0.0119 (0.0031)	0.0071 (0.0016)	0.0264 (0.0042)	0.1043 (0.0236)
7th Story	0.6315 (0.0831)	0.000 (0.0000)	0.4911 (0.0379)	0.0041 (0.0063)	0.0619 (0.0317)	0.000 (0.0000)	0.0634 (0.0124)	0.0001 (0.0002)

Table 6.12: Mean predicted values of final fractional stiffness and strength loss for updating using data from the Northridge earthquake. Standard deviations for simulation results are shown in parentheses.

almost immediately upon loading, since there is no other mechanism for dissipating energy in the model class. Further studies are needed to determine if another form of the viscous-damping matrix is appropriate, or if an alternative source of small-amplitude damping should be used.

6.6 Conclusions

Identification of hysteretic models of structures using earthquake response records is typically an ill-conditioned problem. Bayesian updating using stochastic simulation has the capacity to perform updating of hysteretic structural models, including models with degrading strength and stiffness. This identification approach allows for efficient extraction of relevant information from the data despite the inherent ill-conditioning of the inverse problem involved in system identification.

Bayesian model-class selection was also successfully performed. The results illustrated the usefulness of the information-theoretic interpretation of the evidence; model class complexity is not merely a matter of the number of uncertain parameters, but rather relates to the amount of information extracted from the data.

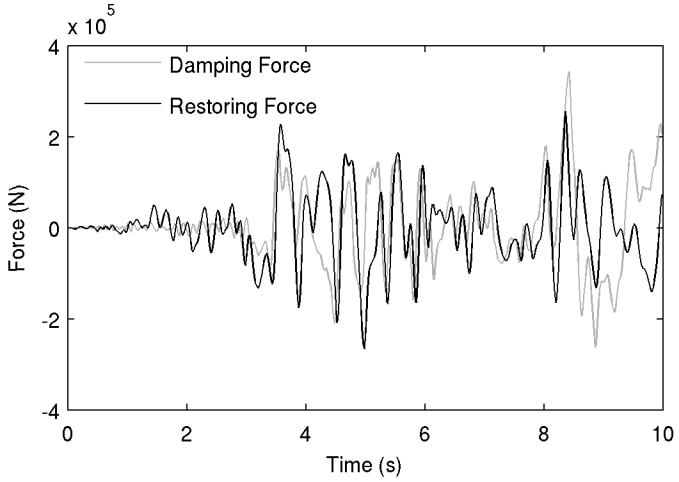


Figure 6.34: The restoring and damping forces acting on the second-floor mass predicted from a model from model class $\mathcal{M}_3$.

Future areas of work might include extending the class of Masing hysteretic models to capture other types of hysteretic behavior, such as "pinching" hysteresis. This type of hysteretic behavior, exhibited in the cyclic loading curve pictured in Figure 6.35, is a common feature in the response of woodframe buildings. It seems possible that by adding an element that captures this type of behavior to the underlying distributed element model, expressions for the restoring forces could be obtained for use with the extended Masing rules.

Another issue that needs further study is the use of viscous damping in hysteretic models to represent small-amplitude energy dissipation. The simulated example problems demonstrated that although the viscous damping forces are relatively small compared to the restoring forces, the exclusion of viscous damping from the identification model can significantly alter the identified hysteretic structural parameters if viscous (i.e., rate-dependent) damping is actually present in the structure. Preliminary results from the study of the Van Nuys hotel suggest that the inclusion of viscous damping may introduce large spurious forces into the calculated structural responses despite apparently reasonable identified modal damping ratios. Though linear viscous damping is perhaps the most commonly used method for introducing small-amplitude energy dissipation into structural models, it is not clear that such rate-dependent dissipation mechanisms are really present in civil structures. Also, the coefficients for

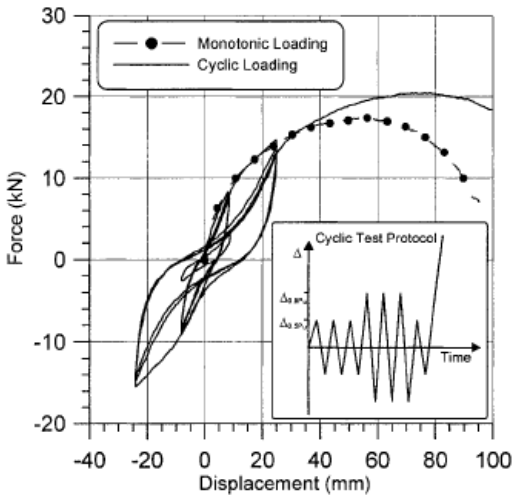


Figure 6.35: Cyclic loading curve for a woodframe shear wall exhibiting "pinched" hysteretic behavior. Taken from Folz and Filiatrault (2001)

a viscous-damping model to give an energy-dissipation mechanism are very difficult to establish from structural drawings during design. In fact, the most likely source of small-amplitude energy dissipation is friction from non-structural elements. These elements could be well modeled by hysteretic elements, suggesting future directions for research.

The final goal of the research presented in this chapter is the use of stochastic simulation techniques to identify Masing hysteretic models for real structures. The initial results, using relatively sparse data from the Van Nuys hotel, identify models that predict behavior that is consistent with observed damage, but this work is still in its preliminary stages. Further investigation into these records and others obtained from buildings that have experienced damage during strong shaking is needed.

Chapter 7

Conclusions and Future Work

The overall goal of the research presented in this dissertation is to demonstrate that stochastic simulation methods are an effective tool for performing parameteric system identification using a Bayesian probabilistic framework. Unlike traditional system identification, where the goal is to find some "optimal" set of parameters, the goal of the Bayesian probabilistic approach is to determine the probability density function over the parameters that, using the theory of probability logic, expresses the plausibility of each set of parameters in the parameter space, conditioned on both prior knowledge and available data. This probabilistic treatment of the problem of system identification allows for the quantification of the uncertainty associated with the identified model and with the predictions made using that model. Additionally, extension of this concept to the model-class level allows for a Bayesian probabilistic approach to model-class selection, which is shown to automatically enforce the concept of "model parsimony."

While Bayesian updating approaches have been applied before to problems of system identification and model-class selection, these methods mostly rely on asymptotic approximations that are difficult to implement in ill-conditioned problems. The use of stochastic simulation methods for Bayesian system identification and model-class selection is a relatively new concept that offers a potential solution for ill-conditioned updating problems. This is important because it is exactly these problems where the Bayesian viewpoint can offer significant advantages over traditional system-identification methods. Issues such as multiple local maxima or ill conditioning, which

are treated as obstacles to be overcome in any type of "optimization" approach to system identification, are viewed in the Bayesian interpretation as statements on the limitations of the information available in the data. These limitations, which are manifested as multiple high-probability regions or manifolds of nearly optimal solutions, can then be accounted for in predictive modeling, instead of being discarded as mere "sub-optimal models," or by selecting only one of multiple optimal models to make predictions.

Chapter 2 presents the Bayesian probabilistic framework for system identification and model-class selection. The methodology of embedding a deterministic model within a probability model is used, and the link between Bayesian updating and typical optimization approaches such as least squares is demonstrated. The process of Bayesian model-class selection is presented, and an information-theoretic interpretation is given in Section 2.2.1, demonstrating that model class complexity can be viewed in terms of the information gained from the data. This interpretation can greatly aid the user's conceptual understanding of the problem.

Chapter 3 introduces the stochastic simulation methods used in this work, showing how they are applied to model updating and to model-class selection, an area where these methods are not commonly applied due to technical challenges.

The example of empirical equations for ground-motion attenuation, presented in Chapter 4, demonstrates the consistency of the results for system identification and model-class selection obtained using two different stochastic simulation methods with results obtained using previously studied Bayesian probabilistic approaches based on asymptotic approximations, an important step in developing confidence in these techniques.

Chapter 5, dealing with the updating of linear structural models with modal data for structural health monitoring, shows how stochastic simulation methods may be applied to extract probabilistic descriptions for the existance and location of damage using ambient response measurements. The simple 2-DOF system presented in Section 5.3 provides a demonstration of the characterization of unidentifiable systems using samples generated with stochastic simulation and shows how model-class se-

lection results may be altered by the choice of prior PDF. Section 5.4 shows that stochastic simulation techniques may be successfully applied to systems with hundreds of uncertain parameters.

The identification performed in Chapter 6 of degrading and non-degrading hysteretic systems demonstrates the potential for stochastic simulation methods to deal with ill-conditioned systems. The class of Masing hysteretic models used are relatively simple, but can capture complex behavior and have an underlying physical basis, preventing unphysical behavior such as excessive drifting. However, the identification of these models using simulated data in Sections 6.2 and 6.4 demonstrates that optimization approaches are difficult to apply, with results that are very sensitive to initial conditions. However, the posterior PDFs of these models can be effectively characterized by samples generated using stochastic simulation. In particular, the identification of deteriorating hysteretic structural models for multi-degree-of-freedom systems using only acceleration responses has not, to the author's best knowledge, been accomplished previously, even for simulated data. In addition, the results of model-class selection provide a more sophisticated example of how model complexity may be viewed in terms of information gained from the data. Finally, the preliminary results using real data from a structure damaged in the 1994 Northridge earthquake are encouraging, in that identified models in one model class exhibit behavior consistent with observed damage, but this data set deserves further study.

Goals for future work should first and foremost include the application of these techniques to ill-conditioned problems involving real data. New stochastic simulation methods, such as the hybrid TMCMC-GS method used in Section 6.5, should be developed and refined with the focus on continually improving computational efficiency. Additionally, the issue of locally identifiable systems, briefly discussed in Section 5.3.2, remains a challenge, even for more advanced stochastic simulation methods, and relates to the problem of clustering, which is relevant in many other fields.

In conclusion, it is believed that ill conditioning is an unavoidable issue in performing system identification on real-world civil structures. A Bayesian probabilistic approach to these problems allows for the ill conditioning to be explicitly considered

instead of being avoided by making unfounded assumptions, and stochastic simulation methods appear to be a promising tool for implementing this approach.

Bibliography

Aagaard, B. T. (1999). *Finite-element simulations of earthquakes*. PhD thesis, California Institute of Technology, Pasadena, California. Also available as California Institute of Technology Report No. EERL-99-03, http://resolver.caltech.edu/CaltechEERL:1999.EERL-99-03.

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, 19:716–723.

Anagnostopoulos, S. A., Roesset, J. M., and Biggs, J. M. (1972). Non-linear dynamic response and ductility requirements of building structures subjected to earthquakes. Technical Report R72-54, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Ashrafi, S. A., Smyth, A. W., and Betti, R. (2005). A parametric identification scheme for non-deteriorating and deteriorating non-linear hysteretic behavior. *Structural Control and Health Monitoring*, 13:108–131.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418.

Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika*, 45:296–315.

Beck, J. L. (1978). *Determining models of structures from earthquake records*. PhD thesis, California Institute of Technology, Pasadena, California. Also

available as California Institute of Technology Report No. EERL-78-01, http://resolver.caltech.edu/CaltechEERL:1978.EERL-78-01.

Beck, J. L. (1989). Statistical system identification of structures. In *Fifth International Conference on Structural Safety and Reliability*, San Francisco, California.

Beck, J. L. (1996). System identification methods applied to measured seismic response. In *Eleventh World Conference on Earthquake Engineering*, Acapulco, Mexico.

Beck, J. L. (2005). personal communication.

Beck, J. L. and Au, S.-K. (2002). Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation. *Journal of Engineering Mechanics*, 128:380–391.

Beck, J. L., Au, S.-K., and Vanik, M. W. (2001). Monitoring structural health using a probabilistic measure. *Computer-Aided Civil and Infrastructure Engineering*, 16:1–11.

Beck, J. L. and Katafygiotis, L. S. (1991). Updating of a model and its uncertainties utilizing dynamic test data. In *First International Conference on Computational Stochastic Mechanics*, Corfu, Greece.

Beck, J. L. and Katafygiotis, L. S. (1998). Updating models and their uncertainties - Bayesian statistical framework. *Journal of Engineering Mechanics*, 124:455–461.

Beck, J. L., Porter, K. A., Shaikhutdinov, R., Au, S.-K., Moroi, T., Tsukada, Y., and Masuda, M. (2002). Impact of seismic risk on lifetime property values, final report, CUREe-Kajima Joint Research Program Phase IV. Technical report, Consortium of Universities for Research in Earthquake Engineering, Richmond, California. http://resolver.caltech.edu/CaltechEERL:2002.EERL-2002-04 .

Beck, J. L. and Yuen, K.-V. (2004). Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics*, 130:192–203.

Benedettini, F., Cappechi, D., and Vestroni, F. (1995). Identification of hysteretic oscillators under earthquake loading by nonparametric models. *Journal of Engineering Mechanics*, 121:606–612.

Bernal, D., Dyke, S. J., Beck, J. L., and Lam, H.-F. (2002). Determination of modal parameters from ambient vibration data for structural health monitoring. In *15th Engineering Mechanics Conference of the American Society of Civil Engineers*, New York, New York.

Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. Wiley.

Boore, D. M., Joyner, W. B., and Fumal, T. E. (1993). Estimation of response spectra and peak accelerations from western North American earthquakes; an interim report. Technical Report 93-509, USGS Open-File Report, Menlo Park, CA.

Boore, D. M., Joyner, W. B., and Fumal, T. E. (1997). Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work. *Seismological Research Letters*, 68:128–153.

Box, G. E. P. and Jenkins, G. (1970). *Time series analysis, forecasting and control*. Holden-Day.

Caicedo, J., Catbas, N., Gul, M., and Zaurin, R. (2006). Benchmark problem for bridge health monitoring: Definition paper. In *4th World Conference on Structural Control and Monitoring*, San Diego, California.

Chiang, D.-Y. (1992). *Parsimonious Modeling of Inelastic Systems*. PhD thesis, California Institute of Technology, Pasadena, California. Also available as California Institute of Technology Report No. EERL-92-02, http://resolver.caltech.edu/CaltechEERL:1992.EERL-92-02.

Chiang, D.-Y. (1999). The generalized Masing models for deteriorating hysteresis and cyclic plasticity. *Applied Mathematical Modelling*, 23:847–863.

Ching, J. (2006). personal communication.

Ching, J. and Beck, J. L. (2004). Bayesian analysis of the Phase II IASC-ASCE structural health monitoring experimental benchmark data. *Journal of Engineering Mechanics*, 130:1233–1244.

Ching, J., Beck, J. L., Porter, K. A., and Shaikhutdinov, R. (2004). Real-time Bayesian state estimation of uncertain dynamical systems. Technical Report EERL 2004-01, California Institute of Technology, Pasadena, California. http://resolver.caltech.edu/CaltechEERL:EERL-2004-01 .

Ching, J. and Chen, Y.-J. (2006). Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. Submitted for publication.

Ching, J., Muto, M., and Beck, J. L. (2005). Bayesian linear structural model updating using Gibbs sampler with modal data. In *Ninth International Conference on Structural Safety and Reliability*, Rome, Italy.

Ching, J., Muto, M., and Beck, J. L. (2006). Structural model updating and health monitoring with incomplete modal data using Gibbs sampler. *Computer-Aided Civil and Infrastructure Engineering*, 21:242–257.

Cifuentes, A. O. and Iwan, W. D. (1989). Nonlinear system identification based on modeling of restoring force behavior. *Soil Dynamics and Earthquake Engineering*, 8:2–8.

Clinton, J. F., Bradford, S. C., Heaton, T. H., and Favela, J. (2006). The observed wander of the natural frequencies in a structure. *Bulletin of the Seismological Society of America*, 96:237–257.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. Wiley.

Cox, R. T. (1961). *The Algebra of Probable Inference*. Johns Hopkins Press.

Cua, G. (2005). *Creating the Virtual Seismologist: developments in ground motion characterization and seismic early warning*. PhD thesis, California Institute of Technology, Pasadena, California. Also available as California Institute of Technology Report No. EERL-2004-12, http://resolver.caltech.edu/CaltechEERL:EERL-2004-12.

Doebling, S. W., Farrar, C. R., Prime, M. B., and Shevitz, D. W. (1996). Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review. Technical Report LA-13070-MS, Los Alamos National Laboratory, Los Alamos, New Mexico.

Eykhoff, P. (1974). *System Identification: Parameter and State Estimation*. Wiley.

Fisher, R. A. (1925). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society*, 22:700–725.

Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. Springer.

Folz, B. and Filiatrault, A. (2001). Cyclic analysis of wood shear walls. *Journal of Structural Engineering*, 127:433–441.

Foutch, D. A., Roeder, C. W., and Goel, S. C. (1986). Preliminary report on seismic testing of a full-scale six story steel building. Technical Report VICU-ENG-86-2009, Universtiy of Illinois, Urbana-Champaign, Urbana, Illinois.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85:972–985.

Gelman, A. B. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457–472.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and

the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

Goodwin, C. G. and Payne, R. L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press.

Grasso, V. F. and Allen, R. M. (2005). Earthquake warning systems: Characterizing prediction uncertainty. In *American Geophysical Union Fall Meeting*, San Francisco, California.

Grasso, V. F., Beck, J. L., and Manfredi, G. (2005). Seismic early warning systems: Procedure for automated decision making. Technical Report EERL 2005-02, California Institute of Technology, Pasadena, California. http://resolver.caltech.edu/CaltechEERL:EERL-2005-02 .

Grigorou, M., Cornell, C., and Veneziano, D. (1979). Probabilistic modeling as decision making. *Journal of Engineering Mechanics*, 105:585–596.

Grünwald, P. (2005). Minumum description length tutorial. In Gruünwald, P., Myung, I. J., and Pitt, M. A., editors, *Advances in Minimum Description Length: Theory and Applications*. MIT Press.

Gull, S. F. (1989). Bayesian inductive inference and maximum entropy. In Skilling, J., editor, *Maximum Entropy and Bayesian Methods*. Kluwer.

Hall, J. F. (2006). Problems encountered from the use (or misuse) of Rayleigh damping. *Earthquake Engineering and Structural Dynamics*, 35:525–545.

Hart, G. C. and Yao, J. T. P. (1976). System identification in structural dynamics. *Journal of Engineering Mechanics*, 103:1089–1104.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57:97–109.

Islam, M. S. (1996). Holiday Inn. In *1994 Northridge Earthquake Buildings Case Study Project Project Proposition 122: Product 3.2*, Technical Report CSSC 94-06. California Seismic Safety Commission, Sacramento, California.

Iwan, W. D. (1966). A distributed-element model for hysteresis and its steady-state dynamic response. *Journal of Applied Mecahnics*, 33:893–900.

Iwan, W. D. (1967). On a class of models for the yielding behavior of continuous or complete systems. *Journal of Applied Mecahnics*, 34:612–617.

Iwan, W. D. and Cifuentes, A. O. (1986). A model for system identification of degrading structures. *International Journal of Earthquake Engineering and Structural Dynamics*, 14:877–890.

Jayakumar, P. (1987). *Modeling and Identification in Structural Dynamics*. PhD thesis, California Institute of Technology, Pasadena, California. Also available as California Institute of Technology Report No. EERL-87-01, http://resolver.caltech.edu/CaltechEERL:1987.EERL-87-01.

Jayakumar, P. and Beck, J. L. (1988). System identification using nonlinear structural models. In Natke, H. G. and Yao, J. T. P., editors, *Structural Safety Evaluation Based on System Identification Approaches*. Vieweg.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106:620–630.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, 4:227–241.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press.

Jeffreys, H. (1932). On the theory of errors and least-squares. *Proceeding of the Royal Society, A*, 138:48–55.

Jeffreys, H. (1939). *Theory of Probability*. Clarendon Press.

Jennings, P. C. (1971). Engineering feautres of the San Fernando earthquake of February 9, 1971. Technical Report EERL 71-02, California Institute of Technology, Pasadena, California. http://resolver.caltech.edu/CaltechEERL:1971.EERL-71-02 .

Joyner, W. B. and Boore, D. M. (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake. *Bulletin of the Seismological Society of America*, 71:2011–2038.

Katafygiotis, L. S. and Lam, H.-F. (2002). Tangential-projection algorithm for manifold representation in unidentifiable model updating problems. *Earthquake Engineering and Structural Mechanics*, 31:791–812.

Kozachenko, L. F. and Leonenko, N. N. (1987). Sample estimates of entropy of a random entropy. *Problems of Information Transmission*, 23:95–101.

Krishnan, S., Ji, C., Komatitsch, D., and Tromp, J. (2006). Case studies of damage to tall steel moment-frame buildings in southern California during large San Andreas earthquakes. *Bulletin of the Seismological Society of America*, 96:1523–1537.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86.

Lai, S.-S. P. and Vanmarcke, E. H. (1980). Overall safety assessment of multistory steel buildings subjected to earthquake loads. Technical Report R80-26, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Lam, H.-F., Yuen, K.-V., and Beck, J. L. (2006). Structural health monitoring via measured Ritz vectors utilizing artificial neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 21:232–241.

Laplace, P. S. (1781). Memoirs sur les probabilités. *Memoirs de l'Acadmie royale des Sciences de Paris*, 9:384–485.

Laplace, P. S. (1812). *Théorie Analytique des Probabilités*. Courcier Imprimeur.

Laplace, P. S. (1951). *A Philosopical Essay on Probabilities; translated from the 6th French ed. by Frederick Wilson Truscott and Frederick Lincoln Emory*. Dover.

Li, Y. R. and Jirsa, J. O. (1998). Nonlinear analyses of an instrumented structure damaged in the 1994 Northridge Earthquake. *Earthquake Spectra*, 14:245–264.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice-Hall, 2nd edition.

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation*, 4:415–447.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

May, B. S. and Beck, J. L. (1998). Probabilistic control for the active-mass-driver benchmark model. *Earthquake Engineering and Structural Dynamics*, 27:1331–1346.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953). Equations of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Muto, M. and Beck, J. L. (2006). Bayesian updating of hysteretic structural models using stochastic simulation. *Journal of Vibration and Control*. (in review).

Natke, H. G. and Yao, J. T. P., editors (1988). *Proceedings of the Workshop on Structural Safety Evaluation Based on System Identification Approaches*. Viewig and Sons.

Neal, R. M. (1993). Probabilistic inference using Markov Chain Monte Carlo methods. Technical report, University of Toronto, Toronto, Canada.

Nelles, O. (2001). *Nonlinear System Identification: from Classical Approaches to Neural Networks and Fuzzy Models*. Springer.

Nielsen, N. N. (1964). *Dynamic responses of multistory buildings*. PhD thesis, California Institute of Technology, Pasadena, California. Also available as California Institute of Technology Report No. EERL.1964.004, http://resolver.caltech.edu/CaltechEERL:1964.EERL.1964.004.

Oh, C. K. and Beck, J. L. (2006). Sparse Bayesian learning for structural health monitoring. In *4th World Conference on Structural Control and Monitoring*, San Diego, California.

Papadimitriou, C., Beck, J. L., and Katafygiotis, L. S. (2001). Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics*, 16:103–113.

Pilkey, W. D. and Cohen, R., editors (1972). *System Identification of Vibrating Structures*. ASME.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14:465–471.

Rissanen, J. (2005). Complexity and information in modeling. In Velupillai, K. V., editor, *Complexity and Constructivity in Economic Analysis*. Blackwell Publishings.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer.

Scholl, R. E., Kutsu, O., Perry, C. L., and Zanetti, J. M. (1982). Seismic damage assessment for high-rise buildings. Technical Report URS/JAB 8020, URS/John A. Blume & Associates, Engineers, San Francisco, California.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Scruggs, J. T., Taflanidis, A. A., and Beck, J. L. (2006). Reliability-based control optimization for active base-isolation systems. *Journal of Structural Control and Monitoring*, 13:705–723.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656.

Sibilio, E., Beck, J. L., Ciampoli, M., and Muto, M. (2006). Bayesian model updating approach for ground-motion attenuation relations. In *ECCM-2006 Conference on Computational Stochastic Structural and Uncertainty Analysis*, Lisbon, Portugal.

Silverman, B. W. (1986). *Density Estimatation for Statistics and Data Analysis*. Chapman and Hall.

Sivia, D. S. (1996). *Data Analysis: A Bayesian Tutorial*. Oxford Science Publications.

Tipping, M. E. (2000). The relevance vector machine. *Advances in Neural Information Processing Systems*, 12:652–658.

Trifunac, M. D., Ivanović, S. S., and Torodovska, M. I. (1999). Instrumented 7-storey reinforced concrete building in Van Nuys, California: Description of the damage from the 1994 Northridge Earthquake and strong motion data. Technical Report CE 99-02, University of Southern California, Los Angeles, California.

Vanik, M. W., Beck, J. L., and Au, S.-K. (2000). A Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics*, 126:738–745.

Wen, Y.-K. (1976). Method for random vibration of hysteretic systems. *Journal of Engineering Mechanics*, 102:249–263.

Yuen, K.-V., Au, S.-K., and Beck, J. L. (2004). Two-stage structural health-monitoring approach for Phase I benchmark studies. *Journal of Engineering Mechanics*, 130:738–745.

Yuen, K.-V. and Beck, J. L. (2003a). Reliability-based robust optimal control for uncertain dynamical systems using feedback of noisy measurements. *Earthquake Engineering and Structural Dynamics*, 32:751–770.

Yuen, K.-V. and Beck, J. L. (2003b). Updating properties of nonlinear dynamical systems with uncertain input. *Journal of Engineering Mechanics*, 129:9–20.

Appendix A

Stochastic Simulation Methods

A.1 Metropolis-Hastings Sampler

Consider a target PDF $f(\theta) \propto p(\mathcal{D} \mid \theta, \mathcal{M})\ p(\theta \mid \mathcal{M})$.

1. Draw the sample parameter vector $\tilde{\theta}_1$ from the prior PDF $p(\theta \mid \mathcal{M})$.

2. Draw a proposed sample parameter vector $\tilde{\theta}_{prop}$ from the proposal PDF $q(\theta \mid \tilde{\theta}_1$.

3. Determine the acceptance ratio r:

$$\begin{aligned} r &= \frac{p(\theta^* \mid \mathcal{D}, \mathcal{M})\ q(\tilde{\theta}_i \mid \theta^*)}{p(\tilde{\theta}_i \mid \mathcal{D}, \mathcal{M})\ q(\theta^* \mid \tilde{\theta}_i)} \\ &= \frac{p(\mathcal{D} \mid \theta^*, \mathcal{M})\ p(\theta^* \mid \mathcal{M})\ q(\tilde{\theta}_i \mid \theta^*)}{p(\mathcal{D} \mid \hat{\theta}_i, \mathcal{M})\ p(\tilde{\theta}_i \mid \mathcal{M})\ q(\theta^* \mid \tilde{\theta}_i)} \end{aligned} \tag{A.1}$$

4. With probability r, set $\tilde{\theta}_2 = \theta_{prop}$, otherwise, set $\tilde{\theta}_2 = \tilde{\theta}_1$.

5. Repeat steps 2 through 4 until the desired number of samples has been simulated.

A.2 Gibbs Sampler

The parameter vector θ is divided into M groups, denoted $\theta_m,\ m = 1, \ldots, M$.

1. Draw samples for all parameter groups except the first, $\tilde{\theta}_m^{(0)},\ m = 2, \ldots, M$ from the prior PDFs $p(\theta_m \mid \mathcal{M})$.

2. Determine the PDF for the first parameter group θ_1 conditioned on the data $\mathcal{D}$ and the samples, $p(\theta_1 \mid \mathcal{D}, \tilde{\theta}_1^{(0)}, \tilde{\theta}_2^{(0)}, \ldots, \tilde{\theta}_M^{(0)}, \mathcal{M})$, and draw the sample $\tilde{\theta}_1^{(1)}$.

A.3 Transitional Markov Chain Monte Carlo

The Transitional Markov Chain Monte Carlo (TMCMC) algorithm, introduced by Ching and Chen (2006) is a method for sampling the posterior PDF of the model class $\mathcal{M}$, defined by the parameter vector θ, the prior PDF over the parameter vector, $p(\theta \mid \mathcal{M})$, and the likelihood function $p(\mathcal{D} \mid \theta, \mathcal{M})$, updated with data $\mathcal{D}$. This is done by estimating a sequence of non-normalized intermediate PDFs, $f_m(\theta)$, $m = 0, \ldots, M$, given by

$$f_m(\theta) \propto p(\mathcal{D} \mid \theta, \mathcal{M})^{\beta_m}\ p(\theta \mid \mathcal{M}) \tag{A.2}$$

where the tempering parameter β_m increases monotonically with m such that $\beta_0 = 0$ and $\beta_M = 1$. The algorithm proceeds as follows:

1. $f_0(\theta) = p(\theta \mid \mathcal{M})$ is chosen by the user and is assumed to be in a form that can be sampled to obtain $\tilde{\theta}_k^{(0)}$, $k = 1, \ldots, N_0$.

2. The value of the tempering parameter for the next level, β_1, is chosen such that the coefficient of variation for $\left\{p(\mathcal{D} \mid \tilde{\theta}_k^{(0)}, \mathcal{M})^{\beta_1 - \beta_0},\ k = 1, \ldots, N_0\right\}$ is equal to some prescribed target value.

3. The plausibility weights for each sample, $w(\tilde{\theta}_k^{(0)}) = p(\mathcal{D} \mid \tilde{\theta}_k^{(0)}, \mathcal{M})^{\beta_1 - \beta_0}$, are calculated for $k = 1, \ldots, N_0$, as is the sample mean S_1 of the N_0 plausibility weights.

4. Samples $\tilde{\theta}_k^{(1)},\ k = 1, \ldots, N_1$ are generated by applying the M-H algorithm as follows: the kth sample is drawn from a Markov chain that starts with the lead sample θ_{lead} equal to one of the samples $\left\{\tilde{\theta}_k^{(0)},\ k = 1, \ldots, N_0\right\}$ where the probability that $\theta_{lead} = \tilde{\theta}_k^{(0)}$ is given by $w(\tilde{\theta}_k^{(0)}) \big/ \sum_{j=1}^{N_0} w(\tilde{\theta}_j^{(0)})$. The M-H algorithm is

applied using a Gaussian proposal PDF that is centered at the current sample in the kth chain; that is, if $\tilde{\theta}_k^{(0)}$ is chosen as the lead sample for the third time, the proposal PDF is centered at the second sample in the chain with $\tilde{\theta}_k^{(0)}$ as the lead. The covariance matrix for the proposal PDF, Σ_m, is given by

$$\Sigma_m = c_m^2 \sum_{k=1}^{N_0} w(\tilde{\theta}_k^{(0)}) \left(\tilde{\theta}_k^{(0)} - \overline{\theta}^{(0)}\right) \left(\tilde{\theta}_k^{(0)} - \overline{\theta}^{(0)}\right)^T \tag{A.3}$$

where

$$\overline{\theta}^{(0)} = \sum_{i=1}^{N_0} w(\tilde{\theta}_i^{(0)}) \tilde{\theta}_i^{(0)} \Big/ \sum_{j=1}^{N_0} w(\tilde{\theta}_j^{(0)}) \tag{A.4}$$

and c_m^2 is a control parameter that is chosen to balance the potential for large MCMC moves with maintaining a reasonable rejection rate.

5. Steps 2 through 4 are repeated until level M, where $\beta_M = 1$ is achieved.

Appendix B

Conditional Distributions for Linear Systems

Consider a system where the output $Y \in \mathrm{R}^M$ is linearly related to the input $X \in \mathbf{R}^N$ through the matrix $A \in \mathbf{R}^{N \times N}$, plus the prediction error $\epsilon \in \mathbf{R}^M$, that is $Y = AX + \epsilon$.

The prior PDF for X is a Gaussian distribution with mean μ_X and covariance matrix Σ_X. The error terms for each measurement are also Gaussian-distributed, with zero mean and a variance $\sigma_j^2,\ j = 1, \ldots, M$ where the prior PDFs over the error variances are inverse gamma distributions defined by the parameters γ_j and δ_j. The prior PDFs of X and σ_j^2 are assumed to be independent.

Given the data $\hat{Y}$, it is desired to find the conditional distributions for X and the prediction-error variances $\sigma_j^2,\ j = 1, \ldots, M$, $p(X \mid \hat{Y}, \sigma^2)$ and $p(\sigma^2 \mid \hat{Y}, X)$.

By Bayes' Theorem,

$$\begin{aligned} p(X \mid \hat{Y}, \sigma^2, \mathcal{M}) &= \frac{p(\hat{Y} \mid X, \sigma^2, \mathcal{M})\ p(X \mid \mathcal{M})\ p(\sigma^2 \mid \mathcal{M})}{p(\hat{Y} \mid \mathcal{M})} \\ &\propto p(\hat{Y} \mid X, \sigma^2, \mathcal{M})\ p(X \mid \mathcal{M}) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}(\hat{Y} - AX)^T(\hat{Y} - AX) - \frac{1}{2}(X - \mu_X)^T\Sigma_X^{-1}(X - \mu_X)\right) \end{aligned} \tag{B.1}$$

Note that the logarithm of this distribution is quadratic in X, showing that the PDF is Gaussian. Thus, the distribution is completely defined by the conditional mean and covariance matrix.

The conditional mean can be determined by maximizing $\log[p(X \mid \hat{Y}, \sigma^2)]$, as

shown below.

$$\frac{\partial p(X \mid \hat{Y}, \sigma^2)}{\partial X} = 0 \tag{B.2}$$

which implies that

$$\begin{aligned} \frac{1}{\sigma^2}(\hat{Y} - AX)^T A &= (X - \mu_X)^T \Sigma_X^{-1} \\ \frac{1}{\sigma^2} A^T (\hat{Y} - AX) &= \Sigma_X^{-1}(X - \mu_X) \\ \frac{1}{\sigma^2} A^T \hat{Y} + \Sigma_X^{-1}\mu_X &= \left(\frac{1}{\sigma^2} A^T A + \Sigma_X^{-1}\right) X \end{aligned} \tag{B.3}$$

Solving for the mean of the conditional distribution,

$$\begin{aligned} \hat{X} &= \left(\frac{1}{\sigma^2} A^T A + \Sigma_X^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} A^T \hat{Y} + \Sigma_X^{-1}\mu_X\right) \\ &= \mu_X + \left(\frac{1}{\sigma^2} A^T A + \Sigma_X^{-1}\right)^{-1} \left(\frac{1}{\sigma^2} A^T \hat{Y} + \Sigma_X^{-1}\mu_X - \Sigma_X^{-1}\mu_X - \frac{1}{\sigma^2} A^T A \mu_X\right) \\ &= \mu_X + \left(\frac{1}{\sigma^2} A^T A + \Sigma_X^{-1}\right)^{-1} \frac{1}{\sigma^2} A^T (\hat{Y} - A\mu_X) \\ &= \mu_X + \left(I + \Sigma_X \frac{1}{\sigma^2} A^T A\right)^{-1} \Sigma_X \frac{1}{\sigma^2} A^T (\hat{Y} - A\mu_X) \end{aligned} \tag{B.4}$$

Introducing the following lemma,

$$(I + PQ)^{-1} P = P(I + QP)^{-1} \tag{B.5}$$

and defining the matrices P and Q as

$$P = \Sigma_X \frac{1}{\sigma^2} A^T, \; Q = A \tag{B.6}$$

it is found that

$$\hat{X} = \mu_X + \Sigma_X A^T (\sigma^2 I A \Sigma_X A^T)^{-1} (\hat{Y} - A\mu_X) \tag{B.7}$$

The conditional covariance matrix is given by

$$\begin{aligned} \text{cov}(X \mid \hat{Y}, \sigma^2) &= -\left[\nabla_X^2(\log[p(X \mid \hat{Y}, \sigma^2)])\right]^{-1} \\ &= \left(\Sigma_X^{-1} + \frac{1}{\sigma^2}A^T A\right)^{-1} \end{aligned} \tag{B.8}$$

Using the matrix identity in Equation B.5 with the definitions for P and Q given in Equation B.6, the covariance matrix can be found:

$$\begin{aligned} \text{cov}(X \mid \hat{Y}, \sigma^2) &= (I + PQ)^{-1}\Sigma_X \\ &= (I + PQ)^{-1}[(I + PQ) - PQ]\Sigma_X \\ &= \Sigma_X - (I + PQ)^{-1}PQ\Sigma_X \\ &= \Sigma_X - P(I + QP)^{-1}Q\Sigma_X \\ &= \Sigma_X - \Sigma_X A^T(\sigma^2 I + A\Sigma_X A^T)^{-1}A\Sigma_X \end{aligned} \tag{B.9}$$

The conditional PDF for the error variance can be expressed as

$$\begin{aligned} p(\sigma^2 \mid \hat{Y}, X) &= \frac{p(\hat{Y} \mid X, \sigma^2)\ p(\sigma^2)}{p(\hat{Y} \mid X)} \\ &\propto p(\hat{Y} \mid X, \sigma^2)\ p(\sigma^2) \\ &\propto \frac{1}{\sigma^n\ \sigma^{2(\gamma+1)}} \exp\left(-\frac{1}{\sigma^2}(\hat{Y} - AX)^T(\hat{Y} - AX) - \frac{\delta}{\sigma^2}\right) \\ &\propto \frac{1}{\sigma^{2(\gamma+\frac{n}{2}+1)}} \exp\left(-\frac{1}{\sigma^2}\left[\delta + \frac{1}{2}(\hat{Y} - AX)^T(\hat{Y} - AX)\right]\right) \end{aligned} \tag{B.10}$$

which is an inverse gamma distribution, such that

$$p(\sigma^2 \mid \hat{Y}, X) = IG\left(\gamma + \frac{n}{2},\ \delta + \frac{1}{2}\sigma(\hat{Y} - AX)^T(\hat{Y} - AX)\right) \tag{B.11}$$